AMBIGUITIES IN RANGE-BASED SYSTEM IDENTIFICATION

BY

ELAD YARKONY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Yuliy Baryshnikov, Chair
Adjunct Professor P. Scott Carney
Professor Minh N. Do
Associate Professor Maxim Raginsky

ABSTRACT

Blind linear system identification (or recovery) arises in several applications in engineering (e.g. channel equalization, super-resolution, MRI and SAR image formation). This is a special case of a bi-linear inverse problem, and is sometimes equivalent to range-based operator recovery.

The aim of this research is to study the structure of solutions for range-based identification, which is typically an affine or projective variety, and is usually ambiguous (containing more than one element - not identifiable).

Algebraic geometry was utilized to derive a generic range-space based identification algorithm and identifiability test. The properties of irreducible complex varieties were used to derive a numerical identifiability guarantee for complex parametric families. In addition, an alternative approach (of so-called preserving pre-compositions) examined the ambiguity from a non-parametric viewpoint, searching for operations that preserve both the structure of a system as well as its range space.

The established framework and results were then used to determine cases wherein the recovery of sampled multichannel finite impulse response (FIR) configurations, particularly blind sampled deconvolution, is ambiguous.

The last chapter of this work offers some insights about the spatial structure of data eigen-patches, that were used in previous chapters in the process of system identification. Empirical results indicate that those eigen-patches tend to exhibit wave-like shapes, and the sample covariance operator is approximately Toeplitz. A heuristic explanation for those two phenomena is offered with some statistical analysis, which could be further developed later into a complete and rigorous explanation of the observations.

I dedicate this work to my beloved parents, Georgette and Abraham,
to whom I owe everything.

שְׁמַ֗ע בְּנִ֣י מוּסַ֣ר אָבִ֑יךָ
וְאַל־תִּטֹּ֥שׁ תּוֹרַ֖ת אִמֶּֽךָ׃

משלי א ח

ACKNOWLEDGMENTS

I would like to thank my adviser, Prof. Yuliy Baryshnikov, for the interesting discussions, for inspiring creative thought, and for offering his remarkably unique (and often - quite beautiful) insights. He helped me, time and again, translate my vaguest intuition into concrete statements, while sharing his endless knowledge and his admiration for aesthetic results.

I would also like to thank my doctoral committee members, each of whom contributed in his own way to my academic and professional growth. Prof. Minh Do, introduced me to the beautiful theory of wavelets and multiresolution analysis, and I later had the pleasure of working as his teaching assistant. Prof. Maxim Raginsky offered compelling lectures on random processes, and his enlightening approach helped me deepen my understanding of this topic. Prof. P. Scott Carney shared with me the fascinating thought-process of a physicist, from which I learned how to stop worrying about pesky abstract nuances, and focus on substance.

I wish to express my appreciation to Prof. Yoram Bresler, who took me in as his student and supported my research for the first two years of the program.

Lastly, I would like to express my deepest gratitude to the people who supported me throughout this lengthy endeavor: my beloved family and my friends, on campus, in town, across the country, and back in Israel.

This research was partially supported by AFOSR FA9550-10-1-05678 and FA9550-11-1-0216 (Baryshnikov, Yarkony), as well as NSF CCF 10-18789 (Bresler, Yarkony).

TABLE OF CONTENTS

LIST OF NOTATIONS AND SYMBOLS

General Algebra

$\mathbf{k}$	a generic field
A^B	set of all functions from B to A (note that $\mathbf{k}^A$ is a field)
$A \setminus B$	set difference (the notation $A - B$ denotes group action)
$\mathbf{k}[x_1, \ldots, x_n]$	polynomial ring in variables $x_1, \ldots, x_n$ and coefficients in $\mathbf{k}$
$\mathbf{k}[V]$	ring of polynomial functions from V to $\mathbf{k}$ (a coordinate ring whenever V is a variety)
$\mathbb{V}(p_1, \ldots, p_n)$	algebraic variety corresponding to $p_1, \ldots, p_n \in \mathbf{k}[x_1, \ldots, x_D]$
$\mathcal{I}(V)$	ideal of polynomials vanishing on a set V
$\mathtt{Gr}(V, r)$	Grassmanian: the set of all r-dimensional subspaces of V
$A \pm B$	Minkowski addition/difference (dilation)

Families of Linear Operators

$\mathtt{Lin}(U, V)$	space of linear maps from a domain U to a co-domain V
$\mathcal{M}$	operator family, a subset of $\mathtt{Lin}(U, V)$
$\mathbf{M}$	parameter set (usually a variety in $\mathbf{k}^D$)
$\mathbf{m}$	a polynomial parametrization $\mathbf{m} : \mathbf{M} \to \mathcal{M}$
$\bar{r}$	typical rank of a family
$\boldsymbol{\rho}(\theta)$	range of $\mathbf{m}(\theta)$
$\mathbf{e}_{\tilde{V}}$	annihilator map of a space $\tilde{V}$ corresponding to $\mathbf{m}$

Operator Pullback Symmetries

$\mathcal{S}_{\mathcal{M}}(S)$	all $\mathcal{M}$ preserving pre-composition of a single operator $S \in \mathcal{M}$
$\mathcal{S}_{\mathcal{M}}$	all $\mathcal{M}$ preserving pre-compositions

Sampled FIR Systems

$\mathbb{T}$	index group (usually finitely generated Abelian, sometimes a Torus)
ς^{τ}	shift/translation operator $x[t] \mapsto x[t+\tau]$
$\langle I \rangle$	subspace of functions in $\mathbf{k}^{\mathbb{T}}$ with support bounded by $I \subset \mathbb{T}$
$\mathtt{D}_I$	sampling on $I \subset \mathbb{T}$ (restriction $\mathbf{k}^{\mathbb{T}} \to \mathbf{k}^I$)
$\mathtt{P}_I$	sampling on $I \subset \mathbb{T}$ (projection $\mathbf{k}^{\mathbb{T}} \to \langle I \rangle$)

Second-Order Statistics of Patches

$\mathbb{E}(X)$	expectation of X
$\mathcal{E}$	sample index (in $\mathbb{T}$)
V	sample ambient space (usually $\mathbb{C}^{\mathbb{T}} \otimes \mathbb{C}^L$)
$\mu_{\mathcal{E}}$	sample mean
$R_{\mathcal{E}}$	sample autocorrelation
r_x	autocorrelation function
η	Haar measure on the index set $\mathbb{T}$ (when normalized will be denoted $\mathbf{p}$)

CHAPTER 1

INTRODUCTION

Linear System Identification aims at determining an unknown input-output model (possibly parameterized) by its inputs, outputs, and priors about its model. When the input is unknown, the problem is classified as *blind identification*, which is the main topic of this work.

1.1 Motivation: Blind Linear System Identification

Algebraically, signals and system can be regarded as vectors and operators over linear spaces. Let U, V be spaces of input and output signals respectively (over some field $\mathbf{k}$), and let $\mathcal{M} \subset \mathtt{Lin}(U, V)$ be a subset of all linear operators $\mathtt{Lin}(U, V)$ mapping U to V.

Blind Identification and Based on Output Data

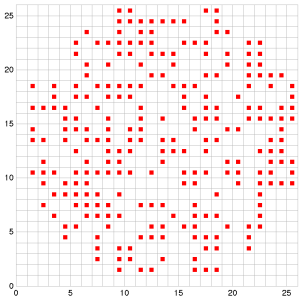
Problem: given a collection of output data $\{y_i\} \subset V$ (here $i \in \mathcal{E}$ is an index), find a linear mapping $S \in \mathcal{M}$ and corresponding inputs $\{x_i\} \subset U$ such that

$$y_i = Sx_i \qquad \text{for all indices } i \in \mathcal{E} \tag{1.1}$$

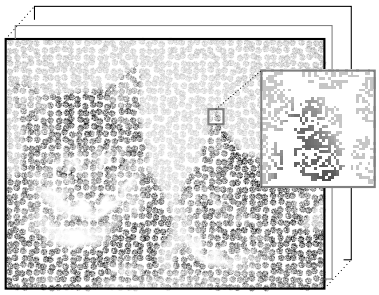
Example 1.1. *An image is blurred by six different kernels, which are then sampled on randomly scattered patches of pixels (Figures 1.1b, 1.1a) on different locations. From this data, the goal is to recover the blurs (system, Figure 1.1c) and possibly the image itself (input, Figure 1.1d).*

We are interested in studying

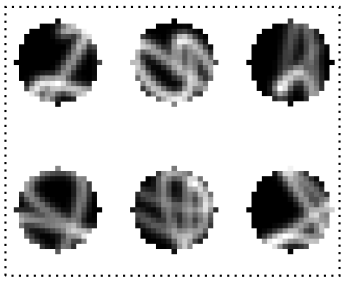
- *How does one recover the blurs and the input?*
- *What sampling patterns enable a unique recovery? (up to scaling)*
- *What blurs have a unique recovery? (up to scaling)*



(a) A patch: only red pixels were sampled



(b) Data ensemble (×6 channels)



(c) System recovery (6 blurs)

(d) Input recovery

Figure 1.1: Blind de-blurring and interpolation

1.1.1 Different Formulations of Blind System Identification

Blind system identification can be formulated as a special case of problems in broader contexts, such as the following:

A Bilinear Inverse Problem

Blind identification is classified as a *bi-linear inverse problem*. In its most general form, a bi-linear problem (see [1] for a general discussion) aims at finding all pairs(a, b) such that

$$c = F(a, b) \tag{1.2}$$

for a bivariate function $F(\cdot, \cdot)$ which is linear in both entries, and some fixed value c. Indeed, the mapping

$$(S, x) \mapsto Sx \ : \ \mathtt{Lin}(U, V) \times U \to V$$

is trivially linear in both S and x. The collection $\mathcal{S}_i = \{(S, x)\} \subset \mathcal{M} \times U$ or all pairs satisfying (1.1) constitute a solution of the bi-linear problem

$Sx = y_i \in V$. There are three solution scenarios, corresponding to what part of the solution is desired:

1. *Joint recovery* if both the input x and system S are sought.
2. *Input recovery* if only the input x is sought.
3. *System recovery* if only the system S is sought.

System identification is concerned with the third case, which is the main focus of this work.

Algebraic Geometry Problem

The bi-linear formalism may not necessarily offer a full and comprehensive description of solution set (not as of writing this work, in any case). It does, however, suggest an important takeaway: a bi-linear solution can be written as a set of *zeros of an algebraic function*:

$$\{(S, x) \mid \mathbf{q}(S, x) = 0\} \qquad \mathbf{q}(S, x) = y - Sx \tag{1.3}$$

The implication of this seemingly trivial formula is that the bi-linear solution (in the finite dimensional case) is *an affine variety* in $\mathtt{Lin}(U, V) \times U^I$, indicating that *algebraic geometry* might be the natural setting for the analysis. The empirical identifiability theory in this work directly stems from properties of complex algebraic varieties.

Tensor Recovery Problem

Recent trends (see Section 1.2 for examples) treat bi-linear problems (specifically emerging from blind channel identification) as *tensor completion problems*. The idea is that for every bi-linear form $F(S, x)$, by conflating S and $x = (x_i) \in U^I$ we can write a multilinear (tensor) form π that agrees with F on rank-1 tensors:

$$F(S, x) = \pi(S \otimes x)$$

For the example, we rewrite $F : \mathbf{k}^{2\times 2} \times \mathbf{k}^2 \to \mathbf{k}^2$ as a multilinear form $\pi : \mathbf{k}^{2\times 2} \otimes \mathbf{k}^2 \to \mathbf{k}^2$:

$$F(S, x) = \begin{bmatrix} S_{11}x_2 - S_{22}x_1 \\ S_{12}x_1 \end{bmatrix} \qquad \rightsquigarrow \pi(W) = \begin{bmatrix} W_{11,2} - W_{22,1} \\ W_{12,1} \end{bmatrix} \tag{1.4}$$

and indeed $\pi(S \otimes x) = F(S, x)$.

In this tensor form (sometimes called "lifting") the inverse bi-linear can be written as follows: given data $y \in V^I$, solve the following for W:

$$\pi(W) = y \in V^I \qquad \text{such that } W \text{ is of rank 1} \tag{1.5}$$

Within the solution set of W, each of which factors to $S \otimes x$, we choose only those that $S \in \mathcal{M}$.

In practice, this is a *matrix completion problem*, wherein $S \otimes x$ is written as $\text{vec}(S)\text{vec}(x)^T$. Think of W as a matrix with missing entries, and complete them such that W if of rank 1, factored as $W = Sx^T$, and $S \in \mathcal{M}$. For the example (1.4) we have

$$W = \begin{bmatrix} W_{11,1} & W_{11,2} \\ W_{12,1} & W_{12,2} \\ W_{21,1} & W_{21,2} \\ W_{22,1} & W_{22,2} \end{bmatrix}$$

If we wish to solve, for instance, $\pi(W) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ then

$$W = \begin{bmatrix} W_{11,1} & \alpha \\ -1 & W_{12,2} \\ W_{21,1} & W_{21,2} \\ \alpha - 1 & W_{22,2} \end{bmatrix}$$

where unresolved entries of W can obtained by ensuring that all 2×2 minors vanish (see similar discussion in [1]), which guarantees that $\text{rk}(W) \leq 1$.

1.1.2 Range Space Based Operator Recovery

If an operator $S \in \mathcal{M}$ maps a set of inputs to outputs on the graph $\{(x_i, y_i)\}_{i \in I}$, namely $y_i = Sx_i$, it does so to any *linear combination* applied simultaneously to the inputs and outputs. For every (finite) scalar collection $\{c_i\}$ one has

$$\sum_i c_i y_i = S\left(\sum_i c_i x_i\right) \tag{1.6}$$

If no restrictions imposed on the input x (allowing linear combinations of $\{x_i\}$ among other things), one can generate many other pairs on the linear

span (usually infinitely more).

In such cases, the relevant information for the recovery of S can be distilled to the output *span* $\text{span}_i\{y_i\}$ rather than the output data itself $\{y_i\}$. Then, the output-data based blind identification problem that was introduced earlier can be equivalently posed for range spaces:

Blind System Identification based on Range Space

Problem: given some subspace $\tilde{V} \subset V$, find $S \in \mathcal{M}$ s.t $\tilde{V} = \texttt{Im}(S)$

Remark 1.2. *The same problem statement applies for kernel-based identification, and the results following are completely analogous.*

Using spans instead of explicit vectors can have practical application; for example, if the outputs $\{y_i\}$ cannot be measured directly, but rather their unknown linear combination (i.e. $\{c_i\}$ in (1.6) are unknown), one can still identify the system S (Lemma 2.20 generalizes that). The idea used in this work is to solve the homogeneous equation

$$\texttt{P}_{\tilde{V}^\perp} S = 0 \text{ subject to } S \in \mathcal{M} \tag{1.7}$$

where $\texttt{P}_H$ is a projection on a space H. This is a generalization of the strategy used in [2], with the exception that $\mathcal{M}$ is not necessarily a linear family.

1.1.3 Solution Existence and Ambiguity

As with any inverse problem, a fundamental question is the *existence* and *uniqueness* (or *ambiguity*) of the solution. For the main concern of this work, the range-based identification (essentially Equation (1.7)) we have:

- **Existence:** a solution $S \in \mathcal{M}$ (1.7) exists only if its range is contained in the given $\tilde{V}$. Strict existence can be relaxed in by approximating $\tilde{V}$ (in practice, converting (1.7) to a minimization problem).
- **Ambiguity:** we have uniqueness only when a given $\tilde{V}$ corresponds to a single $S \in \mathcal{M}$ with $\tilde{V} = \texttt{Im}(S)$. Solutions of blind identification problems notoriously tend to be ambiguous.

Goal Statement

The main goal of this work was to understand the extent to which range-based system identification is ambiguous, or informally, to quantify the *identifiability* of systems.

In the (more general) bi-linear context, a joint solution exists iff both factors exist separately. Thus, existence is equivalent for all solution scenarios (input, system, and joint recovery), but usually not guaranteed - depending on whether the given data is a valid output of some system in $\mathcal{M}$. As for ambiguity, a unique factor does not mean the other is non-ambiguous. For example, given $\{(S, x)\}$, where S is unique with a nontrivial kernel in U, then x is ambiguous with $x + \texttt{Ker}(S)$.

In the tensor formalism, existence of a rank-1 tensor $W = S \otimes x$ does not imply that either S or x satisfies any further restrictions. A solution to the matrix completion problem is merely just one first step in the solution process.

In the special case of homogeneous $\mathcal{M}$ (i.e. $\lambda\mathcal{M} \subseteq \mathcal{M}$ for all scalars λ), the *scaling ambiguity* is inherent to the solution, as every scalar $\alpha \neq 0$ yields

$$(\alpha S)(\frac{1}{\alpha}x) = Sx$$

Note that the tensor formalism $S \otimes x$ eludes this homogeneity ambiguity.

1.2 Related Works

A common application of blind system identification arises in *blind deconvolution* (featured in Example 1.1 and in Chapter 4), aiming at factoring a signal y as a convolution of $\mathbf{h}$ and x (i.e. $y = \mathbf{h} * x$), that are *unknown* factors. *Blind Finite Impulse Response (FIR) deconvolution* further asserts that the filter $\mathbf{h}$ is *finitely supported*. The problem gets more complicated when y is only a partial sampling of the convolution, namely, for some sampling operator $\texttt{D}$ the data is

$$y = \texttt{D}(\mathbf{h} * x)$$

Such models have been utilizes in several applications, such as:

- *Channel equalization*: the reversal of a distortion incurred by a communication channel $\mathbf{h}$, in order to recover the transmitted signal [2,3].

- *Image de-blurring* is a 2D analogous problem of the channel equalization, where the goal is to recover an input image x by a set of blurred measurements y (see [4, 5]). Additional constraints on the blurs are often imposed such as spatial priors (sparsity and gradients), as well as positivity (see e.g. [6]).

- *Super-resolution* (see e.g. [7]): here the output y corresponds to a lower resolution data emerging from a high-resolution image x. The model of S is typically well known, so in that context the solution focuses on the input x.

Recent trends in blind identification employ the inverse bi-linear and matrix completion approach, as described in Section 1.1.1. Several works came up with conditions on the sampling function π under which a tensor can essentially be uniquely recovered [8–10]. The conditions are derived by algebraic and combinatorial considerations. Another related work aims at learning conditions under which a range space can be uniquely associated with a partially sampled matrix [11]. That is, given $\tilde{V} = \texttt{Im}(S)$ and the sample $\pi(S)$ of some unknown S, different completions of $\pi(S)$ to S yield various ranges, but under some conditions on π, the range of S can be determined by $\pi(S)$ solely.

Sparsity priors for bi-linear inverse problems were also incorporated, often relaxing the rank condition to a nuclear norm[1] [9, 10, 12]. This relaxation translates the identification into a convex minimization problem:

$$\min_{W} \|W\|_* \quad \text{subject to } \pi(W) = y$$

More related to this work, algebraic geometry has been playing an important role in the analysis in this research. The applications of some basic tools of algebraic geometry to the questions of identifiability have been undertaken earlier (see e.g. [13, 14]). Those papers address similar questions to the one discussed in this research, with essentially the same toolbox. Their results, however, do not rely on the range (or kernel) identifiability, and do not exploit generic properties, which lie in the foundation of the numerical identifiability test obtained in this research (Section 2.3).

Lastly, the notion of system identification in this work is different, though remotely related to the state-space system identification studied in [15, 16].

[1] Given by the sum of the singular values.

1.3 Established Contributions and Document Outline

The identification problem poses several interesting questions:

1. How can one recover an operator through its range?
2. Under what conditions is an operator identifiable by its range?
3. If one operator is identifiable, does it mean that others are too?
4. How can one test for identifiability of one or more operators?
5. Can increasing the number of channels (i.e. identify several systems simultaneously assuming common input) promote identifiability?
6. Can a system still be identified through partial measurements?

Our goal in this work, is to answer those questions. Some of the early results in this work (e.g. Theorem 2.22) were already outlined in [17], though without the proofs or any detailed analysis.

The analysis can be divided into two main categories, based on the description of the operator family $\mathcal{M}$:

- *Explicit parametrization* of $\mathcal{M}$, where the ambiguity is studied in the parameter space, which is the main topic of Chapter 2.
- *Implicit description* of $\mathcal{M}$ (i.e. algebraic variety), where the range ambiguity is studied in $\mathcal{M}$ itself, discussed in Chapter 3.

The two approaches can be combined and complement each other, as demonstrated in Chapter 4 with the analysis of discrete FIR identifiability.

The following list highlights the contributions of this research and the structure of this document:

1. Formulation of the identification problem in terms of algebraic varieties, and the range-based identification algorithm (Chapter 2).
2. A theorem stating that identifiability is a *generic* property (for complex polynomial parametric models), providing a numerical identifiability guarantee (Section 2.3).
3. Characterization of solutions through *$\mathcal{M}$-preserving pre-compositions* (Chapter 3), using the fact that $\mathtt{Im}(S\mathbf{A}) \subseteq \mathtt{Im}(S)$ for every linear endomorphism $\mathbf{A}$. The introduced algebraic structures characterize

the solution ambiguity in both the operator and parameter space, and can given insights about a potential of a family to be identified through modifications (such as multichannel extensions).

4. Classification of *identifiable* sampled FIR configurations on generalized (discrete, Abelian) index groups (Chapter 4). Traditional FIR identifiability theory has been addressing *one-dimensional sequences or digital images*, and fairly simple (rectangular or periodic) sampling models, often employing the range-space base approach [2,4,7]. One of the interesting conclusions is that uniform sampling of the index group results with an ambiguous identification (Section 4.2.1). Lastly, using pre-composition approach, a combinatorial analysis was developed to test whenever a sampling configuration is not identifiable (Section 4.4).

5. Chapter 5 is a separate endeavor that began while studying the principal components of sample autocorrelations (which stemmed from the parametric identification algorithm, relying on the data span). Examining the way patches occupy an operator's output space (through principal component analysis) revealed an interesting pattern: there are significant wave-like components in natural images, and the ones of lower energy (i.e. constant and "slow" varying) look very much like solutions of a harmonic problem on the patch domain. We take a closer look at this phenomenon, and offer some heuristic explanations.

Some technical notes such as mathematical formulation and proofs are included in the appendices.

CHAPTER 2

IDENTIFIABILITY OF PARAMETRIC OPERATOR FAMILIES

Fix a linear domain and co-domain U and V respectively over a field $\mathbf{k}$. Generally speaking, an operator family can be any subset

$$\mathcal{M} \subset \mathtt{Lin}(U, V) \qquad \text{(Generic Operator Family)}$$

but we will focus on families with structure, specifically algebraic varieties, and parametric models (that are useful in applications in engineering and physics).

This chapter will be structured as follows:

- Section 2.1 features formal definitions of operator families, rank stratification, and specific properties of complex families.
- Section 2.2 discusses the range-based identification of a parametrized operator within a family.
- Section 2.3 highlights one of the main results of this research: a guarantee for identifiability for irreducible complex parametric families (which include linear families).
- Lastly, Section 2.4 concludes with results that are specific for linear families only.

2.1 Algebraic and Parametric Families

We begin with a basic notion of an *operator variety*.

Definition 2.1. *An* operator variety *is an algebraic variety in* $\mathtt{Lin}(U, V)$

$$\begin{gathered} \mathcal{M} = \mathbb{V}(\mathbf{q}_1, \ldots, \mathbf{q}_K) \subset \mathtt{Lin}(U, V) \\ \mathbf{q}_1, \ldots, \mathbf{q}_K \in \mathbf{k}[\mathtt{Lin}(U, V)] \end{gathered} \tag{2.1}$$

where $\{\mathbf{q}_k\}$ *are polynomial mappings taking operators in* $\mathtt{Lin}(U, V)$ *to* $\mathbf{k}$.

This definition gives an implicit description of the operator family, in which $S \in \mathcal{M}$ iff $\mathbf{q}_k(S) = 0$ for all $1 \leq k \leq K$ (or $\mathbf{q}(S) = 0$ in short). We say that $\mathcal{M}$ is a linear family if it is a linear subspace of $\mathtt{Lin}(U, V)$.

Example 2.2. *The set of all rigid transformations on the 2D real plane (in homogeneous coordinates) is an operator variety in* $\mathtt{Lin}(U, V) = \mathbb{R}^{3\times 3}$. *Given* $S = (s_{ij}) \in \mathtt{Lin}(U, V)$, *define the family* $\mathcal{M} = \mathbb{V}(\mathbf{q}_1, \dots \mathbf{q}_6)$ *as the zeros of*

$$\begin{aligned} \mathbf{q}_1(S) &= s_{11} - s_{22} & \mathbf{q}_2(S) &= s_{21} + s_{12} \\ \mathbf{q}_3(S) &= s_{11}s_{22} - s_{21}s_{12} - 1 & \mathbf{q}_4(S) &= s_{23} - 1 \\ \mathbf{q}_5(S) &= s_{31} & \mathbf{q}_6(S) &= s_{32} \end{aligned}$$

This is the set of all 3×3 *real matrices with the structure*

$$\left[\begin{array}{cc|c} q_1 & q_2 & t_1 \\ -q_2 & q_1 & t_2 \\ \hline 0 & 0 & 1 \end{array}\right] \text{ subject to } q_1^2 + q_2^2 = 1$$

An alternative way to describe operator families is by explicit *parametrization*. A *parametric family* of operators is generally a mapping

$$\mathbf{m} : \mathbf{M} \to \mathtt{Lin}(U, V) \tag{2.2}$$

from a *parameter set* $\mathbf{M}$, that is usually embedded in $\mathbf{k}^D$. Every parameter $\theta \in \mathbf{M}$ is associated with some operator $\mathbf{m}(\theta) \in \mathtt{Lin}(U, V)$. An *algebraic family* is a special case of parametric family split out from $\mathtt{Lin}(U, V)$ by *polynomial equations*.

Definition 2.3 (Algebraic Family)**.** *An* algebraic family *is a polynomial image of an algebraic variety in an operator space. That is, the parameter space is an* (affine) algebraic parameter variety*:*

$$\mathbf{M} = \mathbb{V}(p_1, \dots, p_K) \subset \mathbf{k}^D \qquad p_1, \dots, p_K \in \mathbf{k}[X_1, \dots, X_D] \tag{2.3}$$

and a polynomial mapping $\mathbf{m} : \mathbf{k}^D \to \mathtt{Lin}(U, V)$*:*

$$\textit{Operator Family: } \mathcal{M} = \mathbf{m}(\mathbf{M}) \subset \mathtt{Lin}(U, V)$$

By saying that the mapping $\mathbf{m} \in \mathtt{Lin}(U, V)[X_1, \dots, X_D]$ is a polynomial map: the matrix entries of $\mathbf{m}(\theta)$ are *polynomials* in $\theta \in \mathbf{M}$.

Remark 2.4. *Algebraic families make a relaxation of* operator varieties*: every variety is a polynomial image of itself, but not necessarily the other way around (e.g. the projection of the complex variety* $\mathbb{V}(x_1x_2-1)\subset\mathbb{C}^2$ *to* $x_2\in\mathbb{C}$ *is* $\mathbb{C}\setminus\{0\}$ *is not a variety). Only certain types of varieties can be always parametrized (linear families for example).*

All results that concern algebraic families also hold for operator varieties, therefore, analysis will be done on algebraic families when possible.

Remark 2.5. *The collection of all scalar polynomial mappings on an affine variety* $\mathbb{V}$ *is known as* coordinate ring*, denoted* $\mathbf{k}[\mathbb{V}]$*. For an algebraic family* $\mathbf{m}:\mathbf{M}\to\mathtt{Lin}(U,V)$*, the parametrization* $\mathbf{m}$ *is a function in* $\mathbf{k}[\mathbf{M}]\otimes\mathtt{Lin}(U,V)$ *(the collection of all* $\mathtt{Lin}(U,V)$*-valued polynomial functions on* $\mathbf{M}$*).*

Example 2.6. *Let* $\mathbf{M}=U=V=\mathbf{k}^n$*. Define the family* $\mathbf{m}$ *by circulant matrices (on the standard basis):*

$$\mathbf{m}(\theta)=\begin{bmatrix}\theta_0 & & & \theta_2 & \theta_1\\ \theta_1 & \theta_0 & & & \theta_2\\ & & \ddots & \ddots & \\ \theta_{n-1} & & & \theta_1 & \theta_0\end{bmatrix}$$

Example 2.7. *For* $\theta\in\mathbf{M}=\mathbf{k}^D$ *define the autoregressive equation*

$$y[n]=\sum_{i=1}^{D-1}\theta_i y[n-i]+\theta_0 x[n]\qquad\qquad\forall n\geq 0\qquad\qquad(2.4)$$

in x *and* y*, subject to zero-initial condition* $y[n]=0$ *for all* $n<0$*. Given*

$$\mathbf{x}[n]=\begin{bmatrix}x[0] & \dots & x[n]\end{bmatrix}^T\qquad\qquad\mathbf{y}[n]=\begin{bmatrix}y[0] & \dots & y[n]\end{bmatrix}^T$$

one can write $\mathbf{y}[n]=A_n(\theta)\mathbf{x}[n]$ *for some* $(n+1)\times(n+1)$ *matrix* $A_n(\theta)$ *which is polynomial in* θ*. To construct* $A_n(\theta)$ *write* (2.4) *as*

$$y[n]=\bar{\theta}_nA_{n-1}(\theta)\mathbf{x}[n-1]+\theta_0x[0]=\left[\begin{array}{c|c}\bar{\theta}_nA_{n-1}(\theta) & \theta_0\end{array}\right]\mathbf{x}[n]$$

Here $\bar{\theta}_n$ *is a vector of* n *entries associated with* θ *as follows:*

$$\bar{\theta}_n=\begin{bmatrix}0 & \dots & 0 & \theta_D & \dots & \theta_2 & \theta_1\end{bmatrix}\in\mathbb{C}^n$$

and truncated whenever $n < D$. Then $A_n(\theta)$ has the recursive structure:

$$A_n(\theta) = \left[\begin{array}{c|c} A_{n-1}(\theta) & 0 \\ \hline \bar{\theta}_n A_{n-1}(\theta) & \theta_0 \end{array}\right] = \begin{bmatrix} a_0(\theta) & 0 & & 0 \\ a_1(\theta) & a_0(\theta) & & \vdots \\ & & \ddots & \\ a_n(\theta) & a_{n-1}(\theta) & \dots & a_0(\theta) \end{bmatrix} \tag{2.5}$$

of a lower-triangular Toeplitz matrix. The polynomial $a_n(\theta)$ is defined by the recursion $a_0(\theta) = \theta_0$ and

$$a_n(\theta) = \bar{\theta}_n \begin{bmatrix} a_0(\theta) & \dots & a_{n-1}(\theta) \end{bmatrix}^T \tag{2.6}$$

2.1.1 Homogeneous Families and Projectivization

Homogeneous varieties are invariant under scaling. Those objects are central to this study, since linear spaces are homogeneous.

Definition 2.8. *A family $\mathcal{M}$ is* homogeneous *if $\alpha\mathcal{M} \subset \mathcal{M}$ for all $\alpha \in \mathbf{k}$ (the same definition applies specifically to operator varieties). A parametric algebraic family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}$ is said to be* homogeneous *if $\mathbf{M}$ is homogeneous and in addition, $\mathbf{m}$ is also homogeneous, i.e. $\mathbf{m}(\lambda\theta) = \lambda^d \mathbf{m}(\theta)$ for all $\lambda \in \mathbf{k},\ \theta \in \mathbf{M}$.*

Homogeneous families require special treatment in range-identifiability analysis, since range-spaces of operators do not alter under rescaling. This ambiguity is reflected back to the parameter space: $\theta \in \mathbf{M}$ and $\lambda\theta \in \mathbf{M}$ are indistinguishable through observation of the range, as $\mathtt{Im}(\mathbf{m}(\lambda\theta)) = \mathtt{Im}(\mathbf{m}(\theta))$ for any nonzero $\lambda \in \mathbf{k}^*$. This inherent ambiguity is mitigated by replacing $\mathbf{M}$ with its projectivization:

$$\mathbb{P}\mathbf{M} := (\mathbf{M} \backslash \{\mathbf{0}\}) \big/ (\theta \sim \lambda\theta, \lambda \in \mathbf{k}^*) \tag{2.7}$$

In general, when $\mathbf{M}$ is an affine algebraic variety defined by a system of homogeneous polynomials of the same degree d, the projectivization results in a *projective algebraic variety* $\mathbb{P}\mathbf{M}$, and if $\mathbf{M} = \mathbf{k}^D$ then $\mathbb{P}\mathbf{M}$ is the projective space $\mathbb{P}^{D-1}(\mathbf{k})$.

2.1.2 Constructions over Operator Families

Operator families can be combined together to create other families. Alternatively, operator families can sometimes be described as combinations of simpler families.

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two operator families (not necessarily in the same ambient space). An *outer construction* defines new operators by pairs $(S_1, S_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ in various way, depending on whether $\mathcal{M}_1$ and $\mathcal{M}_2$ share the same domain U, co-domain V, or are separate altogether. We consider two cases:

1. Common domain U, that is, $\mathcal{M}_1 \subset \mathtt{Lin}(U, V_1)$ and $\mathcal{M}_2 \subset \mathtt{Lin}(U, V_2)$. The product $\mathcal{M}_1 \times \mathcal{M}_2$ has a natural structure in $\mathtt{Lin}(U, V_1 \oplus V_2)$ defined by
$$(S_1, S_2)x := (S_1 x, S_2 x)$$
and is naturally embedded in $\mathtt{Lin}(U, V_1) \oplus \mathtt{Lin}(U, V_2)$ (much like stacking matrices one over the other). We do not require $\mathcal{M}_1$ and $\mathcal{M}_2$ to be identical.

2. Multichannel extension of *the same model*: multiple instances of the *same* model operate on a *common input* (Figure 2.1). In this case $S \in \mathcal{M}^L$ is the channel stack
$$S = \begin{bmatrix} S_1, & \dots, & S_L \end{bmatrix} : U \to V^L$$
where $S_k \in \mathcal{M}$ corresponds to the k-th channel. Alternative notations are $S \in \mathtt{Lin}(U, V^L)$, $S \in \mathtt{Lin}(U, V \otimes \mathbf{k}^L)$, and with abuse of notations $S \in (\mathtt{Lin}(U, V))^L$ or $S \in \mathtt{Lin}(U, V) \otimes \mathbf{k}^L$.

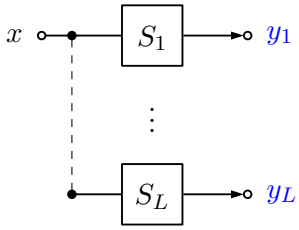


Figure 2.1: Multichannel construction

Constructions over operator families maintain their type: products of operator varieties are operator varieties, products of algebraic families are algebraic, and products of homogeneous families are homogeneous.

Internal constructions are defined between families in the *same* ambient

space $\mathtt{Lin}(U, V)$. If $\mathcal{M}_1, \mathcal{M}_2 \subset \mathtt{Lin}(U, V)$ are two families, the minimal linear family containing both families is the span:

$$\operatorname{span}\{\mathcal{M}_1, \mathcal{M}_2\} = \operatorname{span}\{\mathcal{M}_1\} + \operatorname{span}\{\mathcal{M}_2\}$$

Whenever $\mathcal{M}_1, \mathcal{M}_2$ are linear spaces, then the above is the linear space $\mathcal{M}_1 + \mathcal{M}_2$. Those are used later in Chapter 3 to determine how multichannel extensions affect identifiability.

2.1.3 Rank Stratification of Operator Families

The rank function

$$\operatorname{rk} : \mathtt{Lin}(U, V) \to \mathbb{N}$$

assigns a linear operator with the dimension of its image over $\mathbf{k}$, which plays a significant role in range-identifiability within $\mathcal{M}$ (i.e. a lower-dimensional output space might result with ambiguous operator solutions).

It is often useful to *stratify* the space $\mathtt{Lin}(U, V)$ into classes of different operator ranks. Define

$$\mathtt{Lin}_{\leq r}(U, V) = \{A \in \mathtt{Lin}(U, V) : \operatorname{rk}(A) \leq r\}$$

consisting of operators of rank *at most* r, and likewise $\mathtt{Lin}_{<r}$, $\mathtt{Lin}_{=r}$ etc. The sets $\mathtt{Lin}_{\leq r}$ are *algebraic varieties* (called *determinantal varieties*): for fixed bases in U and V, the set $\mathtt{Lin}_{\leq r}$ is given by the vanishing of all minors of orders $(r+1) \times (r+1)$ in the corresponding matrices (hence the name).

These subsets are nested and define a natural stratification of $\mathtt{Lin}$:

$$0 = \mathtt{Lin}_0(U, V) \subset \mathtt{Lin}_{\leq 1}(U, V) \ldots \subset \mathtt{Lin}_{\leq \min(\mathtt{s},\mathtt{t})}(U, V) = \mathtt{Lin}(U, V)$$

Parametric families $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}(U, V)$ can be stratified in the same way, by introducing:

$$\mathbf{M}_r = \left\{\theta \in \mathbf{M} \;\middle|\; \operatorname{rk}(\mathbf{m}(\theta)) = r\right\}$$

i.e. the set of parameters whose corresponding operator has rank of *exactly* r (similarly $\mathbf{M}_{\leq r}$, where the rank is $\leq r$ etc.). The same hierarchy remains

for the parameter stratification:

$$\mathbf{M}_0 \subset \mathbf{M}_{\leq 1} \subset \mathbf{M}_{\leq 2} \ldots \subset \mathbf{M}$$

The stratification of an operator variety $\mathcal{M} = \mathbb{V}(\mathbf{q})$ can be done by intersection with $\mathtt{Lin}_r$:

$$\mathcal{M}_r := \mathtt{Lin}_r \cap \mathcal{M}$$

likewise $\mathcal{M}_{\leq r}$, where the rank is $\leq r$ etc. Clearly $\mathcal{M}_r = \mathbf{m}(\mathbf{M}_r)$.

2.1.4 Complex Operator Families

In the specific case $\mathbf{k} = \mathbb{C}$ we have three important properties: Euclidean topology, Lebesgue measure, and algebraic closeness. Many results in this work rely on properties of complex varieties that do not necessarily hold with other fields (not even real numbers).

The significance of the complex field is demonstrated in the following example:

Example 2.9. *Consider the family* $\mathbf{m} : \mathbb{C}^3 \to \mathtt{Lin}(\mathbb{C}^2)$ *given by*

$$\mathbf{m}(\theta_1, \theta_2, \theta_3) := \begin{bmatrix} \theta_1 & \theta_3 \\ 0 & \theta_2 \end{bmatrix}$$

The rank of this family equals to 2 *almost everywhere, and drops almost nowhere:*

$$\begin{aligned} \mathtt{Lin}_2 &= \{(\theta_1, \theta_2, \theta_3) \mid \theta_1, \theta_2 \neq 0\} \subset \mathbb{C}^3 \\ \mathtt{Lin}_{\leq 1} &= \mathbb{V}(\theta_1\theta_2) = \{(\theta_1, \theta_2, \theta_3) \mid \theta_1\theta_2 = 0, \} \\ \mathtt{Lin}_0 &= \{0\} \end{aligned}$$

See Figure 2.2 for an illustration on the real space (red shaded planes have rank 1*, the origin, in blue, has rank* 0*). In terms of topological dimensions,* $\mathtt{Lin}_2$ *is a complex manifold of dimension* 3*, while* $\mathtt{Lin}_1$ *is of dimension* 2 *(the union of the planes* $\theta_1 = 0$ *and* $\theta_2 = 0$*) and* $\mathtt{Lin}_0$ *of dimension* 0 *- being a single point in* $\mathbb{C}^3$*.*

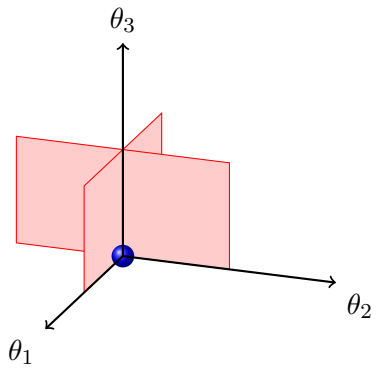


Figure 2.2: Rank strata (of real parameters)

This example illustrates two interesting features:

- The subset $\mathtt{Lin}_{\leq r}$ is nowhere dense in $\mathtt{Lin}_{\leq r+1}$ (and is a nullset).
- The rank function is lower-semicontinuous: it can only increase or remain the same under small perturbations, but not drop. Figuratively speaking, the variety $\mathtt{Lin}_{\leq r-1}$ is a very thin membrane within $\mathtt{Lin}_{\leq r}$.

In general, each *open stratum* consisting of all operators of rank *exactly* r,

$$\mathtt{Lin}_r := \mathtt{Lin}_{\leq r} \setminus \mathtt{Lin}_{\leq r-1}$$

is a *smooth manifold* in $\mathbb{C}^{\mathtt{st}} \cong \mathtt{Lin}(U, V)$ of dimension $r(\mathtt{s} + \mathtt{t}) - r^2$, here $\mathtt{s} = \dim(U)$ and $\mathtt{t} = \dim(V)$.

The following is standard result in algebraic geometry (see [18]):

Proposition 2.10. *Let X be an* irreducible *algebraic variety over $\mathbb{C}$, that is, X cannot be represented as a finite union of algebraic varieties which are not subsets of each other.*[1] *A proper subvariety $Y \subset X$ is nowhere dense in X (or, equivalently, $X \setminus Y$ is everywhere dense on X).*

Informally, it means that imposing any extra polynomial constraint on an irreducible algebraic variety drops the topological dimension of the variety. The irreducibility caveat is required because otherwise, for example, the variety

$$V = \mathbb{V}(x_1x_2, x_1x_3) = \mathbb{V}(x_1) \cup \mathbb{V}(x_2, x_3) \subset \mathbb{C}^3$$

consists of a union of the plane $\mathbb{V}(x_1)$ and the line $\mathbb{V}(x_2, x_3)$ (see Figure 2.3 for real part depiction), whence the former is dense in V and makes a large portion of it.

[1] For example, any linear space is irreducible.

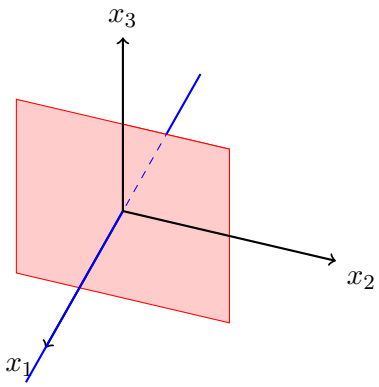


Figure 2.3: A reducible variety $\mathbb{V}(x_1x_2, x_1x_3)$ (illustration in $\mathbb{R}^3$)

One of the implications of Proposition 2.10 is that rank is a generic property:

Corollary 2.11. *Consider an algebraic family* $\mathbf{m} : \mathbf{M} \to \texttt{Lin}$ *where* $\mathbf{M}$ *is* irreducible. *If* $\mathrm{rk}(\mathbf{m}(\theta)) = r$ *for some parameter value* $\theta \in \mathbf{M}$*, then the set* $\mathbf{M}_{<r}$ *(of all parameters where the rank is less than* r*), is a nowhere dense algebraic subvariety of* $\mathbf{M}$*.*

Indeed, $\mathbf{M}_{<r}$ is an algebraic subvariety of the irreducible $\mathbf{M}$ (given by the vanishing minors of matrices with coefficients polynomial in θ), which by Proposition 2.10 and its corollary is nowhere dense in $\mathbf{M}$.

Definition 2.12 (Typical Rank)**.** *We call the largest rank of an operator in the family* $\mathbf{m} : \mathbf{M} \to \texttt{Lin}$ *the* typical rank*, denoted* $\bar{r}$*.*

The Corollary 2.11 implies that if $\mathbf{M}$ is irreducible, then the rank of $\mathbf{m}(\theta)$ is typical *almost everywhere* in $\mathbf{M}$, namely, on a dense subset of $\mathbf{M}$.

The following lemma and its corollary (see proof in Appendix A) is crucial for the theory presented in this chapter, and will be later used to derive the numerical guarantee for identifiability test.

Genericity Lemma

Lemma 2.13. *Let* $X \subset \mathbb{C}^n$ *be a complex irreducible variety, and let* $p : X \to \mathbb{C}^d$ *be some polynomial mapping. Then the set of points in* X *with ambiguous* p*-images,*

$$X_= := \{\theta \mid p(\theta) = p(\tilde{\theta}) \text{ for some } \tilde{\theta} \neq \theta \in X\} \subset X$$

is algebraic*: either dense in* X *or nowhere dense in* X*.*

Corollary 2.14.

1. *The set of points in X with proportional p-images*

$$X_\propto := \{\theta \mid \exists \tilde{\theta} \neq \theta \in X, \quad p(\theta) = \lambda p(\tilde{\theta})\} \subset X$$

   is either dense in X or nowhere dense in X.

2. *The same holds for projective varieties with homogeneous mapping p, with essentially the same proof.*

2.2 Range-based Identification of Algebraic Families

We already learned in Section 1.1.2 that blind system identification (without input constraints) depends on the algebraic span of the data rather than on the data itself, establishing the problem of *range-based identification*.

First an foremost, if $S_1, S_2 : U \to V$ where $\dim(V) \leq \dim(U)$ and both are of full rank, then they cannot be distinguished by their (identical) range. This is a problem especially if the family is typically of full rank. We would normally require that $\dim(V) \leq \dim(U)$ (i.e. a "tall" family). If the family is such that $\dim(V) \leq \dim(U)$, there are several ways to solve it, such as multichannel extensions (as discussed later in Section 3.2), or re-factoring the family when possible (Lemma 2.27).

While in range-based identification the objects of interest are *linear spaces*, for practical purpose, basis vectors are still being used. The statement $\tilde{V} = \boldsymbol{\rho}(\theta) = \mathbf{m}(\theta)U$ materializes computationally with a choice of a basis:

$$U = \operatorname{span}\{u_1, \ldots, u_s\} \quad \Rightarrow \quad \tilde{V} = \operatorname{span}\{\mathbf{m}(\theta)u_1, \ldots, \mathbf{m}(\theta)u_s\}$$

Parameterizing linear spaces by bases is highly ambiguous, being invariant to any full-rank linear combination of the basis elements. We wish to assign V with a *unique set* of parameters that are invariant to bases change (perhaps up to scaling). Fortunately, the Plüker embedding does exactly that, rather elegantly.

2.2.1 Parametrization of Linear Spaces

The *exterior product* (sometimes called wedge product) of vectors generalizes cross-products in $\mathbb{R}^3$. Recall that in $\mathbb{R}^3$, every two-dimensional subspace is spanned by a pair of vectors that are not co-linear, and can be associated with a single normal vector in $\mathbb{R}^3$. The span of two pairs $\{u, v\}, \{u', v'\} \subset \mathbb{R}^3$, can be compared by their cross-product:

$$\text{span}\{u, v\} = \text{span}\{u', v'\} \quad \iff \quad u \times v, u' \times v' \neq 0$$

$$u \times v, u' \times v' \text{are proportional}$$

The exterior product captures the gist of this concept: map any arbitrary ordered basis $\{v_k\}_{k=1}^r$ of $\tilde{U}$ of dimension r to the exterior product

$$\psi : \quad \{v_1, \ldots, v_r\} \mapsto v_1 \wedge \ldots \wedge v_r$$

Under any basis change $\{v_k\} \mapsto \{v'_k\}$ by a matrix A one has

$$\psi(v'_1, \ldots, v'_r) = \det(A)\psi(v_1, \ldots, v_r)$$

In fact, the order of the basis vectors does not matter either, as it at most can flip the sign of the exterior product. As a result, the space $\tilde{U}$ is uniquely associated with the one-dimensional span of $\psi(v_1, \ldots, v_r)$ of some arbitrary basis, which is a class in the projective variety $\mathbb{P}(\bigwedge^r \tilde{U})$. When the vectors $\{v_k\}_{k=1}^r$ are linearly dependent, the exterior product vanish.

The Plüker embedding defines a unique set of *homogeneous* coordinates for range spaces of parametric families. Fix a basis $\{u_k\}_{k=1}^{\mathtt{s}}$ of the domain U, and define the map $\psi : \mathbf{M} \to \bigwedge^{\mathtt{s}}(V)$

$$\psi(\theta) := \bigwedge_{k=1}^{\mathtt{s}} \mathbf{m}(\theta) u_k \tag{2.8}$$

The definition of $\psi(\theta)$ depends on the choice of basis, but a basis change will only scale ψ. Otherwise θ, θ' lead to the same range space if $\psi(\theta), \psi(\tilde{\theta})$ are proportional and nonzero.

It is important to note that if $\mathbf{m}$ is polynomial in θ, then $\psi(\theta)$ is also polynomial in θ. If $\mathbf{m}$ is homogeneous, then ψ is homogeneous as well.

2.2.2 Subspace Observables

Once parameterized by numerical coordinates, range spaces can be thought of as observations in a topological manifold (a *Grassmanian* in this case). In general, when a topological parameter space $\mathbf{M}$ is observed through a mapping

$$\texttt{Obs} : \mathbf{M} \to \mathcal{O}$$

The set $\mathcal{O}$ is called the space of observables, that poses the $\mathcal{O}$-identifiability problem: the extent to which an observation $\texttt{Obs}(\theta)$ identifies a parameter $\theta \in \mathbf{M}$.

Definition 2.15. *A single parameter value θ is said to be* $\texttt{Obs}$-identifiable *if $\texttt{Obs}(\theta) = \texttt{Obs}(\tilde{\theta})$ implies that $\tilde{\theta} = \theta$, i.e. $\texttt{Obs}^{-1}(\texttt{Obs}(\theta)) = \{\theta\}$, is the sole parameter value corresponding to a given observation. Thus, we say that $\theta, \tilde{\theta} \in \mathbf{M}$ are* ambiguous *if $\texttt{Obs}(\theta) = \texttt{Obs}(\tilde{\theta})$, and* non-trivially ambiguous *if $\theta \neq \tilde{\theta}$.*

There is a hierarchy of *parameter identification criteria*:

Definition 2.16. *We say that the parameters of $\mathbf{M}$ are:*

- Globally identifiable *if the mapping* $\texttt{Obs}$ *is one-to-one (an embedding).*
- Almost everywhere identifiable *(or generic identifiability), if there is an open dense subset $\mathbf{M}_o \subset \mathbf{M}$ where* $\texttt{Obs}$ *is identifiable.*
- Locally identifiable at $\theta \in \mathbf{M}$*, if the restriction of* $\texttt{Obs}$ *to some open vicinity of U_θ is identifiable.*

Fix an algebraic family $\mathbf{m} : \mathbf{M} \to \mathcal{M} \subset \texttt{Lin}(U, V)$ of typical rank $\bar{r}$. The image of the *range* mapping $\texttt{Im}$, when restricted to $\texttt{Lin}_{\bar{r}}$, is in the Grassmanian $\texttt{Gr}(V, \bar{r})$, i.e.

$$\texttt{Im} : \texttt{Lin}_{\bar{r}} \to \texttt{Gr}(V, \bar{r}) \qquad\qquad S \mapsto \texttt{Im}(S)$$

Define the range observation $\boldsymbol{\rho} = \texttt{Im} \circ \mathbf{m} : \mathbf{M}_{\bar{r}} \to \texttt{Gr}(V, \bar{r})$:

$$\boldsymbol{\rho}(\theta) := \texttt{Im}(\mathbf{m}(\theta)) \in \texttt{Gr}(V, \bar{r}) \tag{2.9}$$

Range-based parameter identification refers to the identifiability of the observation $\rho : \mathbf{M} \to \mathcal{O} = \texttt{Gr}(V, \bar{r})$, in the terms of Definition 2.16. Thus we will use the notation $\boldsymbol{\rho}$-identifiability to discuss range-identifiability.

Of course, $\boldsymbol{\rho}$ is defined only on the subsets $\mathbf{M}_r$, where the rank is constant, the identifiability question makes sense only on the set of parameters where the rank is typical.

In the homogeneous case, the mapping $\boldsymbol{\rho} : \mathbb{P}\mathbf{M}_{\bar{r}} \to \mathtt{Gr}(V, \bar{r})$ is well defined, and identification aims at finding homogeneous solution, i.e. parameters up to scaling.

2.2.3 Parameter Identification Algorithms

To effectively recover a parameter θ from a range space $\tilde{V} \in \mathtt{Gr}(V, \bar{r})$, we need to write an equation in the parameter θ that includes the range constraint $\tilde{V}$ (the family constraint is inherent to the parametrization, whereas in (1.7) it is an additional constraint).

Let $\tilde{V} \in \mathtt{Gr}(V, \bar{r})$ be an observed range some $\mathbf{m}(\theta) \in \mathtt{Lin}(U, V))$ of typical rank $\bar{r}$. Recall that two subspaces $\tilde{V}, \tilde{V}' \in \mathtt{Gr}(U, \bar{r})$ coincide if their Plücker coordinates are nonzero and proportional, or $\psi(\tilde{V}) \wedge \psi(\tilde{V}') = 0$. Solving the equation below (which is a polynomial equation in θ),

$$A(\theta) = \psi(\tilde{V}) \wedge \psi(\theta) = 0 \tag{2.10}$$

subject to $\psi(\theta) \neq 0$, guarantees that $\boldsymbol{\rho}(\theta) = \tilde{V}$. This equation is not computationally feasible (especially when the involved dimensions are big, the Grassmanian is even bigger).

A slightly more explicit approach utilizes linear projections. Let $\mathbf{q} \in \mathtt{Lin}(V)$ be any linear map that vanishes on $\tilde{V}$. There are infinitely many such maps, but we can explicitly write one of them as the orthogonal complement $\tilde{V}^{\perp} \subset V$:

$$P_{\tilde{V}^{\perp}} = \mathtt{Id} - P_{\tilde{V}} \tag{2.11}$$

assuming that V has some inner product (which is always the case for finite dimensional spaces).

The Annihilator Mapping

The mapping $\mathbf{e}_{\tilde{V}} : \mathbf{M} \to \mathtt{Lin}(U, \tilde{V}^{\perp})$ defined by

$$\mathbf{e}_{\tilde{V}}(\theta) := (\mathtt{Id} - P_{\tilde{V}})\mathbf{m}(\theta) \tag{2.12}$$

is a polynomial mapping in θ. The zeros of this mapping

$$\mathbf{e}_{\tilde{V}}(\theta) = 0 \tag{2.13}$$

correspond to parameters identified by $\tilde{V}$, i.e. $\boldsymbol{\rho}(\theta) \subseteq \tilde{V}$. Therefore, to recover θ, solve (2.13) above. Furthermore, whenever $\mathrm{rk}(\mathbf{m}(\theta))$ is typical, i.e. $\theta \in \mathbf{M}_{\bar{r}}$, then the subspaces coincide: $\boldsymbol{\rho}(\theta) = \tilde{V}$.

Clearly $\mathbf{e}_{\boldsymbol{\rho}(\theta)}(\theta) = 0$ for all θ, but existence of zeros for (2.13) for arbitrary subspace $\tilde{V} \in \mathtt{Gr}(r, V)$ is not guaranteed unless the latter is $\boldsymbol{\rho}(\tilde{\theta})$ for some $\tilde{\theta} \in \mathbf{M}$. Failure to find zeros indicates that the proposed algebraic family does not model the measured data.

Identification Algorithm

Require: measurement coordinates $\{y_\alpha\}$
Require: an algebraic family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}(U, V)$.
  Stack $\{y_\alpha\} \to$ a matrix V
  if V is not of typical rank **then**
    Warning: *non-identifiable parameter or insufficient excitation.*
  end if
  $V \to P_{\tilde{V}}$, a projection onto $\mathrm{colsp}(V)$
  Solve $\mathbf{e}_{\tilde{V}}(\theta) = 0$ for θ.
  if Solution Empty **then**
    Exception: *inadequate model*
  end if
  Return solution(s).

This algorithm does not presume *identifiability* (i.e. unique solution), though can be rather used to *test identifiability* of a given θ, based on the cardinality of the algebraic variety $\mathbb{V}(\mathbf{e}_{\boldsymbol{\rho}(\theta)})$ (we will elaborate on that and introduce a guarantee based on that idea). Of course, when the parametric model is homogeneous (and specifically linear), the solution of $\mathbf{e}_{\tilde{V}}(\theta) = 0$ is also homogeneous and is never unique, so identifiability is pronounced up to scaling, and tested whenever $\mathbb{V}(\mathbf{e}_{\tilde{V}})$ is a one-dimensional line in $\mathbf{M}$.

Example 2.17. *Consider the algebraic (nonlienar) family on* $\mathbf{M} = \mathbb{C}^2$

$$\mathbf{m}(\theta) = \begin{bmatrix} \theta_1 & \theta_2^2 \\ 0 & 2\theta_2 \\ -\theta_2^2 & 0 \\ 0 & 1 \end{bmatrix}$$

We wish to identify θ *such that its associated range is*

$$\tilde{V} = \text{span}\left\{ \begin{bmatrix} 0 \\ -4 \\ -8 \\ -1 \end{bmatrix}, \begin{bmatrix} 10 \\ 8 \\ -4 \\ 2 \end{bmatrix} \right\}$$

The orthogonal $V^\perp$ *spanned by* $\begin{bmatrix} 0 & 1 & 0 & -4 \end{bmatrix}^T$ *and* $\begin{bmatrix} 2 & 0 & 1 & -8 \end{bmatrix}^T$*, therefore the annihilator map,*

$$\mathbf{e}_{\tilde{V}}(\theta) = \begin{bmatrix} 0 & 1 & 0 & -4 \\ 2 & 0 & 1 & -8 \end{bmatrix} \begin{bmatrix} \theta_1 & \theta_2^2 \\ 0 & 2\theta_2 \\ -\theta_2^2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2\theta_2 - 4 \\ 2\theta_1 - \theta_2^2 & 2\theta_2^2 - 8 \end{bmatrix}$$

has a unique zero at $\theta = (2, 2)$*.*

More elaborate examples will be given in Chapter 4.

2.2.4 Ambiguity Structures of Parameters

For a given parameter $\theta \in \mathbf{M}$, a natural question to ask is what (other) parameters in $\mathbf{M}$ are $\boldsymbol{\rho}$-ambiguous with θ. We define the pre-ordering between parameters based the ordering of their respective ranges:

Definition 2.18. *For* $\theta, \tilde{\theta} \in \mathbf{M}$ *define*

$$\tilde{\theta} \preceq \theta \quad : \quad \boldsymbol{\rho}(\tilde{\theta}) \subseteq \boldsymbol{\rho}(\theta) \tag{2.14}$$

and similarly $\prec$ *if the dimension is strictly lower. This ia a pre-order that induces equivalence between parameters:*

$$\tilde{\theta} \sim \theta \quad \Leftrightarrow \quad \boldsymbol{\rho}(\theta) = \boldsymbol{\rho}(\tilde{\theta}) \tag{2.15}$$

which also indicates equal rank of $\mathbf{m}(\theta)$ *and* $\mathbf{m}(\tilde{\theta})$*.*

Equivalent parameters are not necessarily equal, and *not all* pairs in $\mathbf{M}$ are comparable. For a given $\theta \in \mathbf{M}$, we define the set of all parameters bounded by θ with respect to $\preceq$

$$\mathbf{M}_{\preceq\theta} := \{\tilde{\theta} \in \mathbf{M} \mid \tilde{\theta} \preceq \theta\}$$

For every $\theta \in \mathbf{M}$, the set $\mathbf{M}_{\preceq\theta}$ is, in fact, a variety:

Proposition 2.19.

1. $\mathbf{M}_{\preceq\theta} = \mathbb{V}(\mathbf{e}_\theta)$ *here* $\mathbf{e}_\theta := \mathbf{e}_{\rho(\theta)}$ *per* (2.13).
2. *If* $\tilde{\theta} \preceq \theta$ *then* $\mathbb{V}(\mathbf{e}_{\tilde{\theta}}) \subseteq \mathbb{V}(\mathbf{e}_\theta)$, *and* $\theta \sim \tilde{\theta}$ *implies* $\mathbb{V}(\mathbf{e}_\theta) = \mathbb{V}(\mathbf{e}_{\tilde{\theta}})$.

Proof. The first part is trivial by construction of $\mathbf{e}_\theta$. For the second part, let $\phi \in \mathbb{V}(\mathbf{e}_{\tilde{\theta}})$, namely $\mathbf{e}_{\tilde{\theta}}(\phi) = (\mathtt{Id} - P_{\boldsymbol{\rho}(\tilde{\theta})})\mathbf{m}(\phi) = 0$. Since $\tilde{\theta} \preceq \theta$ then

$$\boldsymbol{\rho}(\phi) \subset \mathtt{Ker}(\mathtt{Id} - P_{\boldsymbol{\rho}(\tilde{\theta})}) = \boldsymbol{\rho}(\tilde{\theta}) \subset \boldsymbol{\rho}(\theta) = \mathtt{Ker}(\mathtt{Id} - P_{\boldsymbol{\rho}(\theta)})$$

so that $\phi \in \mathbb{V}(\mathbf{e}_\theta)$ as well. The equality follows from mutual inclusion. □

If $\tilde{\theta} \in \mathbb{V}(\mathbf{e}_\theta)$ then the two parameters are *comparable* with $\theta \preceq \tilde{\theta}$, but not necessarily *equivalent*, for they may have different ranks. The set of all parameters equivalent to θ

$$\{\tilde{\theta} \mid \boldsymbol{\rho}(\theta) = \boldsymbol{\rho}(\tilde{\theta})\} = \mathbb{V}(\mathbf{e}_{\tilde{\theta}}) \cap \mathbf{M}_{\bar{r}}$$

is dense in the variety $\mathbb{V}(\mathbf{e}_{\tilde{\theta}})$.

When $\mathcal{M}$ is given without parametrization (implicitly, as a variety) we define the set of all operators in $\mathcal{M}$ that share the same range as a specific operator $S \in \mathcal{M}$:

$$\mathcal{M}_{\preceq S} := \{S \in \mathcal{M} \mid \boldsymbol{\rho}(T) \subset \boldsymbol{\rho}(S)\}$$

and $\mathcal{M}_S$, $\mathcal{M}_{=S}$ etc.

2.2.5 Rank Deficiency and Sufficient Excitation Condition

Throughout the discussion, we were concerned thus far only with parameters of *typical rank*, i.e. $\mathbf{M}_{\bar{r}}$. Rank deficient parameters in parameters in $\mathbf{M}_{<\bar{r}}$ have been found empirically to be non-identifiable. A general statement

about identifiability of parameters in $\mathbf{M}_{<\bar{r}}$ is an open problem as of the time of writing this document.

In addition, the default assumption that the dimension of the output span indeed meets the typical rank $\bar{r}$

$$\dim \operatorname{span}_i\{y_i\} = \bar{r} \tag{2.16}$$

This can only happen if the input vectors $\{x_i\}$ fully span the input space U, a condition known as *sufficient excitation*. Otherwise a lower dimension $\boldsymbol{\rho}(\theta)$ has larger ortho-complement, thus imposing less constraints on $\tilde{\theta}$ in the annihilator equation $\mathbf{e}_{\boldsymbol{\rho}(\theta)}(\tilde{\theta}) = 0$.

It is important to mention that (2.16) can be broken by insufficient excitation as well as rank-deficient parameter. However, there is no way to tell which of the conditions fail (perhaps both), without additional priors, just like the equation $xy = 0$ cannot specify which factor vanish.

2.2.6 Transformations That Preserve Identifiability

The lemma below (see proof in Appendix A) suggests that range identifiability of algebraic families remains intact under several transformations.

Lemma 2.20. *If θ is a $\boldsymbol{\rho}$-identifiable parameter of an algebraic family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}$, then it is also $\boldsymbol{\rho}$-identifiable in the families $\mathbf{m}_1$, $\mathbf{m}_2$, $\mathbf{m}_3$ defined as follows:*

$$\tilde{\mathbf{m}}_1(\theta) := \mathbf{m}(\theta)S, \qquad S \in \mathtt{Lin}(\tilde{U}, U) \text{ is surjective} \tag{2.17}$$

$$\tilde{\mathbf{m}}_2(\theta) := T\mathbf{m}(\theta), \qquad T \in \mathtt{Lin}(V, \tilde{V}) \text{ is injective} \tag{2.18}$$

$$\tilde{\mathbf{m}}_3(\theta) := \mathbf{m}(f(\theta)), \qquad f : \tilde{\mathbf{M}} \to \mathbf{M} \text{ is invertible} \tag{2.19}$$

Note that $\tilde{\mathbf{m}}_3$ is an algebraic family only if f is a polynomial map.

The Lemma 2.20 carries into local $\boldsymbol{\rho}$-identifiability of the family, almost-everywhere $\boldsymbol{\rho}$-identifiability and global $\boldsymbol{\rho}$-identifiability.

Note that if $\mathbf{M}(\theta)$ has the factorization $T\tilde{\mathbf{M}}(\theta)S$ where S is injective and T is surjective, then $\tilde{\mathbf{M}}$ has the same identifiability as $\mathbf{M}$. In particular, rank-deficient linear families can be reduced to full-rank families that way (see Lemma 2.27).

2.2.7 Rational Families

Most of the theory discussed here applies to *rational* families as well, whose matrix entries (of $\mathbf{m}(\theta)$) for a fixed bases of U and V are rational functions in θ. Equivalently, one can write

$$\mathbf{m}(\theta) = \frac{1}{\mathbf{m}_d(\theta)}\mathbf{m}_n(\theta)$$

Here $\mathbf{m}_d \in \mathbf{k}[\mathbf{M}]$ is the least common multiplier of all the denominators of all entries, and $\mathbf{m}_n : \mathbf{M} \to \mathtt{Lin}(U, V)$ is an algebraic family.

Note that whenever $\mathbf{m}_d(\theta) \neq 0$ then $\mathtt{Im}(\mathbf{m}(\theta)) = \mathtt{Im}(\mathbf{m}_n(\theta))$, thus

$$\mathbf{m}(\theta) \text{ is identifiable} \quad \Longleftrightarrow \quad \mathbf{m}_n(\theta) \text{ is identifiable, and } \mathbf{m}_d(\theta) \neq 0$$

namely, identification of a rational $\mathbf{m}(\theta)$ amounts to identification of algebraic $\mathbf{m}_n(\theta)$, up to avoiding parameters in the subvariety $\mathbb{V}(\mathbf{m}_d)$ where $\mathbf{m}$ is not defined (and is usually nowhere dense in $\mathbf{M}$ anyhow).

Example 2.21. *Let $\theta = (\theta_n, \theta_d) \in \mathbf{k}^N \times \mathbf{k}^D$ define an autoregressive moving-average equation on sequences:*

$$\sum_{i=0}^{D-1} \theta_d[i]y[n-i] = \sum_{i=0}^{N-1} \theta_n[i]x[n-i] \tag{2.20}$$

with initial condition $y[n] = 0$ for all $n < 0$. Then

$$\begin{aligned} y[0] &= \frac{\theta_n[0]}{\theta_d[0]}x[0] \\ y[1] &= \frac{\theta_n[0]}{\theta_d[0]}x[1] + (\frac{\theta_n[1]}{\theta_d[0]} - \frac{\theta_d[1]\theta_n[0]}{\theta_d[0]^2})x[0] \\ &\vdots \end{aligned}$$

By induction $y[n]$ linearly depends on $\mathbf{x}[n] := \Big[x[0], \quad \ldots, \quad x[t]\Big]^T$ and generally we have a linear mapping $\mathbf{y}[n] = A_n(\theta)\mathbf{x}[n]$ where $A_n(\theta)$ is a rational function in θ. This is also an example of a homogeneous family, as

$$\mathbf{m}(\lambda\theta) = \mathbf{m}(\theta)$$

for all nonzero $\lambda \in \mathbf{k}$.

2.3 Identifiability Guarantee for Complex Algebraic Families

We already have a strategy to test the identifiability of a single parameter - out of infinitely many others. A surprising result shows that identifiability is a *generic* property of a complex algebraic family, that is, if one parameter is identifiable, then all of them but a nullset are identifiable as well.

Generic Identifiability of Irreducible Families

Consider some algebraic family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}$ of typical rank $\bar{r}$.

Theorem 2.22. *If* $\mathbf{M}$ *is an irreducible variety, then* $\boldsymbol{\rho}$*-identifiability is a generic property in* $\mathbf{M}$*.*

Proof. This is a direct consequence of Lemma 2.13. Fix an arbitrary basis $\{u_k\} \subset U$, and recall the polynomial mapping (2.8):

$$p(\theta) := \bigwedge_{k=1}^{r} \mathbf{m}(\theta) u_k$$

The collection of all parameters θ with proportional images under p is either dense or nowhere dense in $\mathbf{M}$. □

As a result of Theorem 2.22, *a numerical guarantee* for the range identifiability can be provided by testing *a single parameter value.*

Numerical Identifiability Guarantee

Corollary 2.23. *A* random *parameter value, drawn from any density continuous with respect to the Lebesgue measure on the (open stratum) of* $\mathbf{M}$*, is identifiable with probability* 1 *if and only if the family is almost everywhere identifiable.*

The case study in Chapter 4 utilizes this guarantee for linear models, though we should remark that this guarantee is valid for nonlinear families, which distinguishes this work from previous results.

2.3.1 The Homogeneous Case

In the homogeneous case, one resort to identification *up to a scalar* by replacing the parameter space $\mathbf{M}$ with its projectivized version, $\mathbb{P}\mathbf{M}$ as defined in (2.7). The observable $\boldsymbol{\rho} : \mathbb{P}\mathbf{M}_{\bar{r}} \to \mathtt{Gr}(V, \bar{r})$ is well defined due to

the homogeneity (and $\mathbf{M}_r$ is obviously homogeneous for all r). The same the nomenclature of identifiability persists in the homogeneous case: global, almost everywhere etc., - for this case, even while the rescaling ambiguity is intrinsically there. Identifiability up to scaling in $\mathbf{M}$ is equivalent to identifiability in $\mathbb{P}\mathbf{M}$.

The Theorem 2.22 holds in the homogeneous case as well.

Theorem 2.24. *Let* $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}$ *be a homogeneous algebraic family of typical rank is* $\bar{r}$ *with an irreducible* $\mathbf{M}$. *The corresponding range observable defined on the dense subset of* $\mathbb{P}\mathbf{M}$ *is:*

$$\boldsymbol{\rho} : \mathbb{P}\mathbf{M}_{\bar{r}} \to \mathtt{Gr}(V, \bar{r})$$

If there exists a single parameter value $\theta \in \mathbb{P}\mathbf{M}_{\bar{r}}$ *which is* $\boldsymbol{\rho}$*-(i.e. range)-identifiable, then the family* $\mathbf{m} : \mathbb{P}\mathbf{M} \to \mathbb{P}\mathtt{Lin}$ *is* almost everywhere identifiable *with respect to* $\boldsymbol{\rho}$.

2.4 Results for Linear Families

Linear families are directly related to the problem of blind multichannel deconvolution that motivated this study. Linear structure exhibit homogeneity (which adds the complexity of dealing with with projective spaces), but compensate with a straightforward analysis.

A family $\mathcal{M} \subset \mathtt{Lin}(U, V)$ is said to be linear whenever $\mathcal{M}$ is a linear space. Correspondingly, a parametric family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}(U, V)$ is said to be *linear* whenever $\mathbf{M}$ is a linear space and $\mathbf{m}$ is a linear map. To make the analysis strategy more transparent, it helps to fix the range of $\mathbf{m}$ with a basis, say $\{S_\sigma\} \subset \mathtt{Lin}(U, V)$ (whence $\mathbf{M} = \mathbf{k}^{\mathbf{\Sigma}}$), so that

$$\mathbf{m}(\theta) = \sum_{\sigma \in \mathbf{\Sigma}} \theta_\sigma S_\sigma \tag{2.21}$$

We can make the following assumptions without loss of generality:

1. $\mathbf{M} = \mathbf{k}^{\mathbf{\Sigma}}$, $U = \mathbf{k}^{\mathtt{s}}$ and $V = \mathbf{k}^{\mathtt{t}}$ are *Cartesian spaces*, and the operator space is the matrix space $\mathtt{Lin} = \mathbf{k}^{\mathtt{t} \times \mathtt{s}}$, assuming standard (or any other fixed) bases for U, V. In what follows, the analysis of linear families will be done on matrices rather than operators, that is, $\mathbf{m} : \mathbf{k}^D \to \mathbf{k}^{\mathtt{t} \times \mathtt{s}}$.

2. $\mathbf{m}$ is injective - and has a trivial kernel in $\mathbf{M}$ (otherwise, replace $\mathbf{M}$ with the quotient space $\mathbf{M}/\mathtt{Ker}(\mathbf{m})$ or $\mathtt{Ker}(\mathbf{m})^\perp$, and adjust $\mathbf{m}$ accordingly).

3. $\mathbf{m}$ has full typical rank (otherwise, it can be factored out, see Lemma 2.27).

4. For identifiability, we require that $\mathtt{t} > \mathtt{s}$ (i.e. "tall" matrix family) which implies $\bar{r} = \mathtt{s}$. Otherwise if $\mathtt{t} \leq \mathtt{s}$, then $\boldsymbol{\rho}(\theta) = V$ for almost all $\theta \in \mathbf{M}$.

When the family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}(U, V)$ is linear, then $\mathbf{e}_{\tilde{V}}$ in (2.12) is linear in θ. Now the obvious corollary to Theorem 2.24 reads can be stated in terms of nullity.

Identifiability Testing Criterion for Linear Families

Corollary 2.25. *If the kernel of $\mathbf{e}_{\tilde{\theta}}(\theta)$ is a one-dimensional subspace of $\mathbf{M}$ (and thus generated by $\tilde{\theta}$), then $\tilde{\theta}$ is $\boldsymbol{\rho}$-identifiable, and $\mathbf{m}$ is $\boldsymbol{\rho}$-identifiable almost everywhere. Conversely, if $\mathbf{m}$ is $\boldsymbol{\rho}$-identifiable almost everywhere, then there exists a subset of $\mathbf{M}$ of full (Lebesgue) measure, such that for any $\tilde{\theta}$ in this subset, the kernel of $\mathbf{e}_{\tilde{\theta}}$ is one-dimensional.*

The kernel of $\mathbf{e}_{\tilde{V}}$ is given by the homogeneous linear equation in $\mathbf{k}^{\boldsymbol{\Sigma}}$:

$$\mathbf{e}_{\tilde{V}}(\theta) = (\mathtt{Id} - P_{\tilde{V}})\mathbf{m}(\theta) = \sum_{\sigma \in \boldsymbol{\Sigma}} \theta_\sigma (\mathtt{Id} - P_{\tilde{V}}) S_\sigma = 0 \tag{2.22}$$

which determines identifiability: existence, if (2.22) has a nontrivial solution space, and uniqueness if its nullity is 1.

Example 2.26. *Suppose that we want to test the identifiability of the matrix family $\mathbf{m} : \mathbb{C}^2 \to \mathbb{C}^{4\times 3}$ defined by*

$$\mathbf{m}(\theta) = \begin{bmatrix} \theta_1 & \theta_2 & 0 \\ \theta_1 & 0 & \theta_2 \\ 0 & \theta_1 & \theta_2 \\ \theta_2 & 0 & \theta_1 \end{bmatrix}$$

Throw a random parameter, say $\theta_* = (1, 1)$. The orthogonal to the range is spanned by $\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}^T$ and the annihilator map is

$$\mathbf{e}_{\theta_*}(\theta) = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}^T \begin{bmatrix} \theta_1 & \theta_2 & 0 \\ \theta_1 & 0 & \theta_2 \\ 0 & \theta_1 & \theta_2 \\ \theta_2 & 0 & \theta_1 \end{bmatrix} = \begin{bmatrix} \theta_1 - \theta_2 & 0 & \theta_2 - \theta_1 \end{bmatrix}$$

is a homogeneous equation in θ with a one-dimensional solution spanned by $\operatorname{span}\{(1,1)\}$ *(note that θ_* is in that span). Therefore, the family is range identifiable almost everywhere in* $\mathbf{M}$.

We remark on some necessary conditions for almost everywhere image identifiability for linear $\mathbf{M}$: at the very least, the dimension of the variety $\mathbb{P}\mathbf{M}$ should be at most the dimension of the typical Grassmanian $\mathtt{Gr}(V, \bar{r})$. Recalling that the dimension of $\mathtt{Gr}(V, \bar{r})$ is $(\mathtt{t} - \bar{r})\bar{r}$, where $\mathtt{t} = \dim(V)$, we see that

$$\dim(\mathbb{P}\mathbf{M}) = \dim(\mathbf{M}) - 1 \leq (\mathtt{t} - \bar{r})\bar{r} \tag{2.23}$$

is necessary for range-identifiability of linear families. This condition will be later exploited in Chapter 4 to determine what kind of sampling patterns and how many channels are required for multichannel FIR identification.

2.4.1 Rank Deficient Families

Incidentally, when the typical $\bar{r}$ of a linear family is not full, the rank deficiency can be factored out.

Lemma 2.27. *Let $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}$ be linear with typical rank $\bar{r}$. There exists a factorization*

$$\mathbf{m}(\theta) = T\tilde{\mathbf{m}}(\theta)S^* \tag{2.24}$$

where $\tilde{\mathbf{m}} : \mathbf{M} \to \mathbb{C}^{\tilde{\mathtt{t}} \times \tilde{\mathtt{s}}}$ has full typical rank (i.e. $\min(\tilde{\mathtt{t}}, \tilde{\mathtt{s}}) = \bar{r}$, and S, T of appropriate dimensions are independent on θ).

See proof in Appendix A.

By Lemma 2.20, if S, T have full rank then identifiability of $\mathbf{m}$ is can be tested by the one of the reduced $\tilde{\mathbf{m}}$. This is particularly useful for cases in which the ambient spaces dimensions are much larger than the data.

Note that Lemma 2.27 does not hold for nonlinear families, as in the next example:

Example 2.28. *Consider the family $\mathbf{m} : \mathbb{C}^2 \to \mathbb{C}^{2\times 2}$ defined by*

$$\mathbf{m}(\theta) := \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}$$

Clearly, the typical rank of this family is 1*. Assume by contradiction that it has a factorization of the form* (2.24)*. Then* $\tilde{\mathbf{m}}(\theta)$ *should be a scalar (rank* 1*), and so the operator*

$$S = \mathbf{m}(\theta) = T\tilde{\mathbf{m}}(\theta)S^* = \tilde{\mathbf{m}}(\theta)\underbrace{TS^*}_{C}$$

is a matrix with entries all of the form $S_{i,j} = \tilde{\mathbf{m}}(\theta)C_{i,j}$*. However, note that the entries of*

$$\mathbf{m}(\theta) = \begin{bmatrix} \theta_1^2 & \theta_1\theta_2 \\ \theta_1\theta_2 & \theta_2^2 \end{bmatrix}$$

do not have a common polynomial factor, which is a contradiction.

CHAPTER 3

$\mathcal{M}$-PRESERVING PRE-COMPOSITIONS

So far, we have studied the operator identification ambiguity from the perspective of the parameter space $\mathbf{M}$. Looking at an operator $S \in \mathcal{M}$, if there exists some $\mathbf{A} \in \mathtt{Lin}(U)$ such that $S\mathbf{A} \in \mathcal{M}$ as well, then clearly S cannot be identified in $\mathcal{M}$ by its range, simply because $\mathtt{Im}(S\mathbf{A}) \subseteq \mathtt{Im}(S)$ (see Lemma 3.1). In this chapter, we will characterize the ambiguity of range-identification by studying what operations on $\mathcal{M}$ preserve both its structure and range.

The resulting concept is a useful computational tool to study ambiguity of operator varieties, and linear families especially.

3.1 Range Invariants

Recall a classical result in linear algebra:

Lemma 3.1. *Two operators $S, T \in \mathtt{Lin}(U, V)$ admit $\mathtt{Im}(S) \subseteq \mathtt{Im}(T)$ if and only if there exists $\mathbf{A} \in \mathtt{Lin}(U)$ such that $S = T\mathbf{A}$.*

Consequently, we define *pre-composition* as composing S with $\mathbf{A} \in \mathtt{Lin}(U)$:

Definition 3.2 (Pre-Composition)**.** *Let $S \in \mathtt{Lin}(U, V)$ and $\mathbf{A} \in \mathtt{Lin}(U)$. The* pre-composition *map associated with $\mathbf{A}$ is defined as the linear self-mapping of $\mathtt{Lin}(U, V)$ given by $S \mapsto S\mathbf{A}$.*

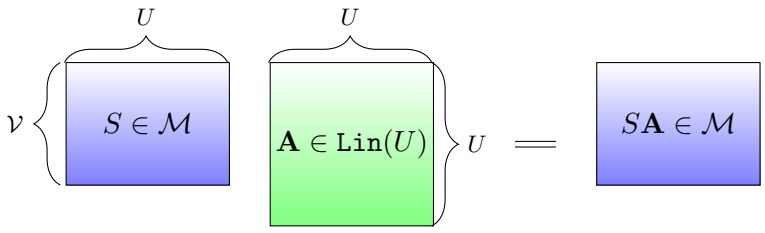


Figure 3.1: $\mathcal{M}$-preserving pre-composition map

By Lemma 3.1, any pre-composition keeps the range intact or shrinks it, $\mathtt{Im}(S\mathbf{A}) \subseteq \mathtt{Im}(S)$ with equality when $\mathbf{A}$ has full rank. We will be interested finding pre-compositions $\mathbf{A}$ of $\mathcal{M}$ that intersect $\mathcal{M}$.

The identity $\mathbf{A} = \mathtt{Id}$ is a trivial example, and $\mathbf{A} = \lambda\mathtt{Id}$ for homogeneous families. Formally we define the following:

Definition 3.3. *Let $\mathcal{M} \subset \mathtt{Lin}(U, V)$. We say that $\mathbf{A} \in \mathtt{Lin}(U)$ is:*

1. Weak $\mathcal{M}$-preserving *with respect to a single $S \in \mathcal{M}$, if $S\mathbf{A} \in \mathcal{M}$.*
2. Weak $\mathcal{M}$-preserving *with respect to a subset $\mathcal{M}' \subset \mathcal{M}$, if $\mathcal{M}'\mathbf{A} \subset \mathcal{M}$.*
3. Strong $\mathcal{M}$-preserving, *if $\mathcal{M}\mathbf{A} \subset \mathcal{M}$.*

The corresponding sets of $\mathcal{M}$-preserving pre-compositions in $\mathtt{Lin}(U)$ shall be defined as

$$\begin{aligned} \mathcal{S}_{\mathcal{M}}(S) &:= \{\mathbf{A} \ : \ S\mathbf{A} \in \mathcal{M}\} \\ \mathcal{S}_{\mathcal{M}}(\mathcal{M}') &:= \{\mathbf{A} \ : \ \mathcal{M}'\mathbf{A} \subset \mathcal{M}\} \\ \mathcal{S}_{\mathcal{M}} &:= \{\mathbf{A} \ : \ \mathcal{M}\mathbf{A} \subset \mathcal{M}\} \end{aligned}$$

There is a clear hierarchy

$$\mathtt{Id} \in \mathcal{S}_{\mathcal{M}} \subset \mathcal{S}_{\mathcal{M}}(S) \tag{3.1}$$

so that $\mathcal{M}$-preserving sets, weak or strong, are never empty. When $\mathcal{S}_{\mathcal{M}}(A) = \{\mathtt{Id}\}$ for $A = S, \mathcal{M}'$ or $\mathcal{M}$ we say that it is *trivial* (the same definition is valid with $\lambda\mathtt{Id}$ for the homogeneous case).

The set $\mathcal{S}_{\mathcal{M}}$ is simple to compute in some cases. One of the main conclusions of this work is that families with nontrivial strong $\mathcal{S}_{\mathcal{M}}$ do not have an identifiable multichannel extension.

3.1.1 Identifiability and $\mathcal{M}$-Preserving Sets

The $\mathcal{M}$-preserving pre-compositions make a useful computational tool in range identifiability: an operator $S \in \mathcal{M}$ is range identifiable if $\mathcal{S}_{\mathcal{M}}(S) = \{\mathtt{Id}\}$ (or $\lambda\mathtt{Id}$ if $\mathcal{M}$ is homogeneous), in which case $\mathcal{S}_{\mathcal{M}}(S)$ is said to be *trivial*.

Corollary 3.4. *The immediate conclusion from* (3.1) *is that a necessary (but insufficient) condition for identifiability is that $\mathcal{S}_{\mathcal{M}}$ is trivial.*

The analogous condition in the homogeneous case is $\mathcal{S}_{\mathcal{M}}(S) = \text{span}\{\texttt{Id}\}$, namely only rescaling of S is preserved in $\mathcal{M}$.

Proposition 3.5. *Let $\mathcal{M}$ be a homogeneous family. For every sub-family $\mathcal{M}' \subset \mathcal{M}$ and every operator $S \in \mathcal{M}'$ we have*

$$\text{span}\{\texttt{Id}\} \subseteq \mathcal{S}_{\mathcal{M}} \subseteq \mathcal{S}_{\mathcal{M}}(\mathcal{M}') \subseteq \mathcal{S}_{\mathcal{M}}(S) \tag{3.2}$$

where span$\{\texttt{Id}\}$ *is the space of all scaler operators.*

See proof in Appendix A.

Before we continue to some examples, it is worth mentioning that pre-compositions act as linear column combinations of matrix representations (obtained once bases are fixed). Searching for $\mathbf{A}$ such that $S\mathbf{A} \in \mathcal{M}$ amounts to finding all column linear combinations, that maintain $SA \in \mathcal{M}$.

The simplest type of operation on columns is moving one column from one index to another (including to itself). All other linear column combinations can be written as a linear combination of such moves.

Example 3.6. *Consider the four-dimensional linear family $\mathcal{M} \subset \mathbf{k}^{3\times 7}$ whose matrix structure is depicted in Figure 3.2. Certain column moves are prohibited (e.g. $3 \to 2$, $5 \to 4$) for they end outside the permitted family structure. Other moves such as $1 \leftrightarrow 2$), $2 \to 3$* might *be allowed, but require further checking: they* entail *moves that might be prohibited for certain matrices in this family. For example, the move $1 \to 2$ entail the moves $3 \to 4$ and $4 \to 5$, all of which are allowed. The move $2 \to 1$ entails the move $4 \to 3$ which is prohibited.*

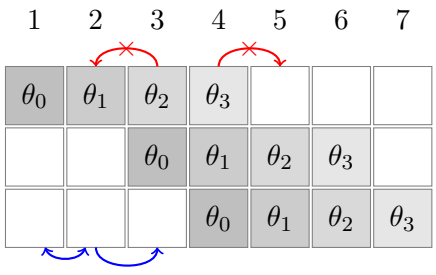


Figure 3.2: Preserving structured matrix by column operations

If we find nontrivial moves that preserve the matrix structure (e.g. $1 \to 2$ in the above example), then this model will never be identifiable - even in a multichannel setup. This kind of combinatorial strategy can be used to

determine when families of certain types are identifiable or not, which is the main topic discussed in Section 3.4.1 and later in Section 4.4.

Example 3.7. *Consider the three-dimensional family in* $\texttt{Lin}(\mathbb{C}^5, \mathbb{C}^2)$:

$$\mathbf{m}(\theta_1, \theta_2, \theta_3) := \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & 0 & 0 \\ 0 & 0 & \theta_1 & \theta_2 & \theta_3 \end{bmatrix}$$

- *Let* $S_1 = \mathbf{m}(1,0,0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$*; for* $\mathbf{A} \in \mathbb{C}^{5\times 5}$ *we have*

$$S_1\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix}$$

  To maintain structure, namely, $S\mathbf{A} = \mathbf{m}(\alpha, \beta, \gamma)$ *for some* α, β, γ*:*

$$\mathbf{A} \in \mathcal{S}_{\mathcal{M}}(S_1) \quad \Rightarrow \quad \mathbf{A} = \begin{bmatrix} \alpha & \beta & \gamma & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ 0 & 0 & \alpha & \beta & \gamma \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}$$

  hence $\mathcal{S}_{\mathcal{M}}(S_1)$ *is linearly parameterized by* 18 *parameters.*

- *For* $S_2 = \mathbf{m}(0,1,0)$ *and* $S_3 = \mathbf{m}(0,0,1)$ *we get* $\mathbf{A}$ *of the forms*

$$\mathcal{S}_{\mathcal{M}}(S_2) = \left\{ \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ \alpha & \beta & \gamma & 0 & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ 0 & 0 & \alpha & \beta & \gamma \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \right\}$$

$$\text{and } \mathcal{S}_{\mathcal{M}}(S_3) = \left\{ \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \alpha & \beta & \gamma & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ 0 & 0 & \alpha & \beta & \gamma \end{bmatrix} \right\}$$

  respectively, which are linear spaces with 18 *degrees of freedom.*

- *For strong preserving* $\mathcal{S}_{\mathcal{M}} = \mathcal{S}_{\mathcal{M}}(S_1) \cap \mathcal{S}_{\mathcal{M}}(S_2) \cap \mathcal{S}_{\mathcal{M}}(S_3)$ *(see Equation*

(3.9) *below) hence we require*

$$\mathbf{A} = \begin{bmatrix} \alpha & 0 & 0 & 0 & 0 \\ \beta & \gamma & \delta & 0 & 0 \\ 0 & 0 & \alpha & 0 & 0 \\ 0 & 0 & \beta & \gamma & \delta \\ 0 & 0 & 0 & 0 & \alpha \end{bmatrix}$$

which is a four-dimensional linear space. This is no coincidence that the degree of ambiguity is the square of the shift between the first and second row of $\mathbf{m}(\theta)$ *(see Theorem 4.10).*

3.1.2 The Structure of $\mathcal{M}$-Invariants

When $\mathcal{M} = \mathbb{V}(\mathbf{q}(S)) \subset \mathtt{Lin}(U, V)$ is an operator variety, it turns out that the various preserving sets $\mathcal{S}_{\mathcal{M}}$ in Definition 3.3 are also algebraic varieties.

The weak preserving set $\mathcal{S}_{\mathcal{M}}(S)$ is a variety in $\mathtt{Lin}(U)$ given by

$$\mathcal{S}_{\mathcal{M}}(S) = \{A \ : \ \mathbf{q}(SA) = 0\} = \mathbb{V}(p) \subset \mathtt{Lin}(U) \tag{3.3}$$

where $p(A) := \mathbf{q}(SA)$. This is an implicit structural constraint on $\mathcal{S}_{\mathcal{M}}(S)$. The strong preserving set can be written as the intersection

$$\mathcal{S}_{\mathcal{M}} = \bigcap_{S \in \mathcal{M}} \mathcal{S}_{\mathcal{M}}(S) \subset \mathtt{Lin}(U) \tag{3.4}$$

This is an intersection of closed sets (in the Zariski topology), which is a closed set on its own. By Hilbert's basis theorem (see [19]), there must be a finite set of polynomials spanning the ideal $\mathcal{I}(\mathcal{S}_{\mathcal{M}})$. In practice, it means that the strong $\mathcal{S}_{\mathcal{M}} = \mathbb{V}(p)$ is an algebraic variety defined by some polynomial mapping $p(\mathbf{A})$.

3.1.3 Relations between $\mathcal{S}_{\mathcal{M}}$ and Ambiguous Parameters

Recall that for an algebraic family $\mathbf{m} : \mathbf{M} \to \mathtt{Lin}(U, V)$, we defined the variety of ambiguous parameters $\mathbf{M}_{\preceq\theta}$. It is natural to ask how this set compares to $\mathcal{S}_{\mathcal{M}}(\mathbf{m}(\theta))$. Indeed, there is a correspondence between the two: define $\mathbf{\Psi}_\theta : \mathbf{M} \to \mathtt{Lin}(U)$ by

$$\mathbf{\Psi}_\theta(\tilde{\theta}) := \mathbf{m}(\theta)^\dagger \mathbf{m}(\tilde{\theta})$$

Here $\mathbf{m}(\theta)^\dagger$ is the Moore-Penrose pseudo-inverse. Clearly $\mathbf{\Psi}_\theta$ is a polynomial mapping in $\tilde{\theta}$ (as $\mathbf{m}$ is).

Lemma 3.8.

1. *$\mathbf{\Psi}_\theta$ maps $\mathbf{M}_{\preceq\theta}$ into $\mathcal{S}_\mathcal{M}(\theta) := \mathcal{S}_\mathcal{M}(\mathbf{m}(\theta))$.*
2. *$\mathbf{\Psi}_\theta$ is injective on $\mathbf{M}_{\preceq\theta}$ in the sense that $\mathbf{\Psi}_\theta(\tilde{\theta}_1) = \mathbf{\Psi}_\theta(\tilde{\theta}_2)$ implies $\mathbf{m}(\theta_1) = \mathbf{m}(\theta_2)$ (and $\theta_1 = \theta_2$ if $\mathbf{m}$ is injective).*

Proof. For every $\tilde{\theta} \in \mathbf{M}_{\preceq\theta}$, we have

$$\begin{aligned}\mathbf{m}(\theta)\mathbf{\Psi}_\theta(\tilde{\theta}) &= \mathbf{m}(\theta)\mathbf{m}(\theta)^\dagger\mathbf{m}(\tilde{\theta}) \\ &= \mathtt{P}_{\boldsymbol{\rho}(\theta)}\mathbf{m}(\tilde{\theta}) && (\boldsymbol{\rho}(\tilde{\theta}) \subset \boldsymbol{\rho}(\tilde{\theta})) \\ &= \mathbf{m}(\tilde{\theta})\end{aligned}$$

so that $\mathbf{\Psi}_\theta(\tilde{\theta}) \in \mathcal{S}_\mathcal{M}(\theta)$.

For the injectivity assume that $\tilde{\theta}_1, \tilde{\theta}_2 \in \mathbf{M}_{\preceq\theta}$ such that $\mathbf{\Psi}_\theta(\tilde{\theta}_1) = \mathbf{\Psi}_\theta(\tilde{\theta}_2)$. By Lemma 3.1 we know that $\mathbf{m}(\theta_1) = \mathbf{m}(\theta)\mathbf{\Psi}_\theta(\tilde{\theta}_1)$ and $\mathbf{m}(\theta_2) = \mathbf{m}(\theta)\mathbf{\Psi}_\theta(\tilde{\theta}_2)$ so the two trivially coincide: $\mathbf{m}(\theta_1) = \mathbf{m}(\theta_2)$. □

Mapping in the opposite direction, i.e. from $\mathcal{S}_\mathcal{M}(\theta)$ to $\mathbf{M}_{\preceq\theta}$, is less obvious, since $\mathbf{m}$ is usually not invertible (and even if so, the inverse is rarely a polynomial). However, the mapping

$$\mathbf{A} \mapsto \mathbf{m}(\theta)\mathbf{A}$$

offers a partial inverse, into $\mathcal{M}$ rather than all the way back into the parameter space $\mathbf{M}_{\preceq\theta}$. Trivially, $\mathbf{m}(\theta)\mathcal{S}_\mathcal{M}(\theta) \mapsto \mathcal{M}$ (by definition of the former), and so $\mathbf{\Psi}_\theta\mathbf{m}(\theta)\mathbf{A} = \mathbf{A}$ when restricted to $\mathbf{A} \in \mathcal{S}_\mathcal{M}(\theta)$.

To conclude, we showed the existence of a polynomial map between $\mathbf{M}_{\preceq\theta}$ to $\mathcal{S}_\mathcal{M}(\theta)$, and a linear map between $\mathcal{S}_\mathcal{M}(\theta)$ back to $\mathcal{M}_{\preceq\theta}$. The two are unfortunately not necessarily isomorphic as algebraic varieties. In the special case of linear families, both $\mathbf{M}_{\preceq\theta}$ and $\mathcal{S}_\mathcal{M}(\theta)$ are linear spaces, and the mapping $\mathbf{A}_\theta$ is a linear isomorphism between the two (see Proposition 3.13 below).

3.2 Preserving Constructions of Families

The goal of this section is to understand how $\mathcal{S}_\mathcal{M}$ (weak or strong) change under linear constructions over $\mathcal{M}$.

Proposition 3.9. *For every collection of families* $\mathcal{M}_1, \ldots, \mathcal{M}_n \subset \mathtt{Lin}(U, V)$*, the product family* $\mathcal{M}_1 \times \ldots \times \mathcal{M}_n$ *containing multichannel maps*

$$S = \Big(S_1, \quad \ldots, \quad S_n\Big) \in \mathtt{Lin}(U, V^n)$$

has the weak preserving set:

$$\mathcal{S}_{\mathcal{M}_1 \times \ldots \times \mathcal{M}_n}(S) = \bigcap_{i=1}^{n} \mathcal{S}_{\mathcal{M}_i}(S_i) \tag{3.5}$$

See proof in Appendix A.

In the special case where $\mathcal{M}_i$ are all copies of the same model, we get

$$\mathcal{S}_{\mathcal{M}^L}(S) = \bigcap_{i=1}^{L} \mathcal{S}_{\mathcal{M}}(S_i) \tag{3.6}$$

The obvious corollary of (3.6) is that the preserving set of the multichannel $\mathcal{M}^L$ is smaller than each of its components. The conclusion is similar to the one suggested previously in the parametric setup: increasing the channel count of a family $\mathcal{M}$ is generally a good idea, and may sometimes decrease the ambiguity of range identification.

Let $\mathcal{M}_1, \ldots, \mathcal{M}_n \subset \mathcal{M}$ be a collection of sub-families of some linear family $\mathcal{M}$. Let $\mathcal{M}'$ be the minimal linear subspace of $\mathcal{M}$ containing all $\mathcal{M}_i$

$$\mathcal{M}' = \operatorname{span}\{\mathcal{M}_1, \ldots, \mathcal{M}_n\} \tag{3.7}$$

which is generated by all linear combinations of representatives from each sub-family. We remark that if $\mathcal{M}_1, \ldots, \mathcal{M}_n$ are all linear families, then $\mathcal{M}' = \mathcal{M}_1 + \ldots + \mathcal{M}_n$.

The set of pre-compositions preserving $\mathcal{M}'$ within $\mathcal{M}$ is easy to describe, as per the following proposition (see proof in Appendix A):

Proposition 3.10. *For* $\mathcal{M}'$ *in* (3.7) *we have*

$$\mathcal{S}_{\mathcal{M}}(\mathcal{M}') = \bigcap_{i=1}^{n} \mathcal{S}_{\mathcal{M}}(\mathcal{M}_i) \tag{3.8}$$

The implication of this proposition will be useful for linear systems, in which studying $\mathcal{S}_{\mathcal{M}}$ will be equivalent to studying $\mathcal{S}_{\mathcal{M}}(S_i)$ for each of the basis elements of $\mathcal{M}$ (see Equation (3.9)).

3.3 $\mathcal{M}$-Preserving in the Linear Case

Assume now that $\mathcal{M}$ is a linear family with the standard parametrization obtained by a basis $\{S_\sigma\}$, where $\mathbf{M} = \mathbf{k}^\Sigma$ and $\mathbf{m}(\theta) = \sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma S_\sigma$.

Proposition 3.11. *Is $\mathcal{M}' \subset \mathcal{M}$ is a linear sub-family, and $S \in \mathcal{M}'$, then*

1. *$\mathcal{S}_\mathcal{M}$, $\mathcal{S}_\mathcal{M}(\mathcal{M}')$, and $\mathcal{S}_\mathcal{M}(S)$ are subspaces of $\mathtt{Lin}(U)$.*
2. *$\mathcal{S}_\mathcal{M}$ is a sub-algebra of $\mathtt{Lin}(U)$, thus $\mathcal{M}$ is a module over $\mathcal{S}_\mathcal{M}$.*

The proof can be found in Appendix A.

Recall that a linear $\mathcal{M}$ can be written as a linear variety in $\mathtt{Lin}(U, V)$, i.e. a zero set of a linear (polynomial) mapping $p : \mathtt{Lin}(U, V) \to \mathbf{k}^d$. In that case, $\mathcal{S}_\mathcal{M}(S)$ is also a linear variety, defined by the zeros of the mapping $\mathbf{q}(\mathbf{A}) = p(S\mathbf{A})$, and for a parametric case $S = \sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma S_\sigma$, this reads

$$\mathbf{q}(\mathbf{A}) = \sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma p(S_\sigma \mathbf{A})$$

The strong $\mathcal{M}$-preserving set can be written as

$$\mathcal{S}_\mathcal{M} = \bigcap_{\sigma\in\mathbf{\Sigma}} \mathcal{S}_\mathcal{M}(S_\sigma) = \bigcap_{S\in\mathcal{M}} \mathcal{S}_\mathcal{M}(S) \tag{3.9}$$

for every basis $\{S_\sigma\}$ of $\mathcal{M}$.

Example 3.12. *Consider the family $\mathbf{m} : \mathbb{C}^{4\times 6} \to \mathbb{C}^{16\times 14}$ defined by embedding blocks as depicted in Figure 3.3. For a random value of θ, the weak preserving $\mathcal{S}_\mathcal{M}(\mathbf{m}(\theta))$ is a linear space spanned by the 14×14 identity matrix $\mathtt{Id}$, and two other matrices (computed empirically) plotted in Figure 3.4.*

3.3.1 Ambiguous Parameters and $\mathcal{M}$-Preserving

We already established the relation between $\mathbf{M}_{\preceq\theta}$ and $\mathcal{S}_\mathcal{M}(\theta)$ by a polynomial mapping. In particular, for a linear family $\mathbf{m}(\theta)$, this correspondence happens to be linear: one can always find a basis for $\mathbf{M}_{\preceq\theta}$ in $\mathbf{M}_{\bar{r}}$, (as the latter is dense in $\mathbf{M}_{\preceq\theta}$), denoted $\{\theta = \theta_0, \ldots, \theta_K\} \subset \mathbf{M}_{\bar{r}}$, and let $\{\mathtt{Id} = \mathbf{A}_0, \ldots, \mathbf{A}_K\}$ be their corresponding preserving pre-compositions, i.e. $\mathbf{m}(\theta_k) = \mathbf{m}(\theta_0)\mathbf{A}_k$.

Proposition 3.13. *There is a linear correspondence between ambiguous parameters in $\mathbf{M}_{\preceq\theta}$ and the pre-compositions in $\mathcal{S}_\mathcal{M}(\theta)$. That is, for every*

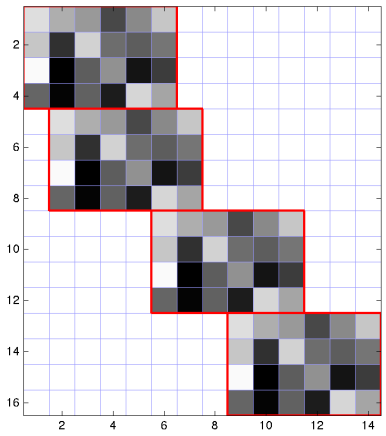


Figure 3.3: Depiction of a matrix family defined by repeated blocks

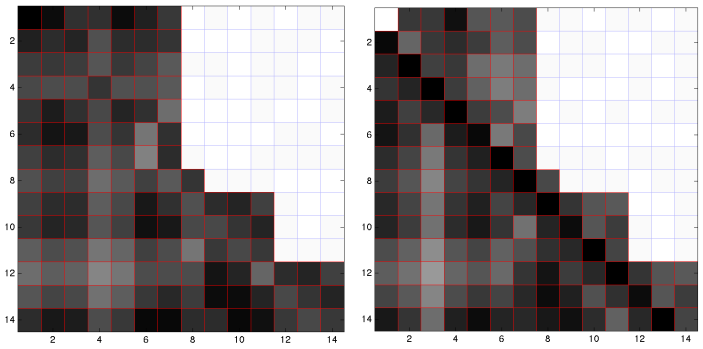


Figure 3.4: Depiction of preserving pre-composition matrices (other than identity)

ambiguous parameter θ *with the coefficients* $\{c_k\}_{k=1}^K$*, the corresponding pre-composition is given by the same coefficients with respect to the basis* $\mathbf{A}_k$*:*

$$\mathbf{m}(\sum_{k=0}^{K} c_k\theta_k) = \mathbf{m}(\theta_0)\sum_{k=0}^{K} c_k\mathbf{A}_k$$

The proof is trivial, due to the linearity of $\mathbf{m}$. Furthermore, if $\mathbf{m}$ is injective on $\mathtt{Lin}(U,V)$, as it certainly should, then $\{\mathbf{A}_k\}_{k=0}^K$ are linearly independent (for every generating $\theta_0 \in \mathbf{M}_{\bar{r}}$), then we have the linear isomorphism $\mathcal{S}_{\mathcal{M}}(\theta) \cong \mathbf{M}_{\preceq\theta}$ (between the two, as vector spaces) given by the mapping above.

3.3.2 Ambiguity in Linear Multichannel Extensions

The preserving structure of a multichannel extension (acting on a common input) is related to the single-channel. For weak preserving sets, we reiterate

the result of (3.6)

$$\mathcal{S}_{\mathcal{M}^I}(S) = \bigcap_{i\in I} \mathcal{S}_{\mathcal{M}}(S_i) \tag{3.10}$$

Here S_i are the different channel components of S. Surprisingly, the strong preserving set of a family, is maintained for tensor extensions as well.

Proposition 3.14. *For every finite I we have $\mathcal{S}_{\mathcal{M}^I} = \mathcal{S}_{\mathcal{M}}$.*

See proof in Appendix A. The relaxation for the case of a direct sum of different families $\mathcal{M}_1, \mathcal{M}_2 \subset \mathtt{Lin}(U, V)$ is the intersection

$$\mathcal{S}_{\mathcal{M}_1 \oplus \mathcal{M}_2} = \mathcal{S}_{\mathcal{M}_1} \cap \mathcal{S}_{\mathcal{M}_2}.$$

It has already been stated that more channels may decrease ambiguity. As it turns out from the Theorem 3.15 below (see proof in Appendix A), there is a bound for channel count, over which increasing has no effect, and the intersection in (3.10) reaches its lowest possible limit - which is $\mathcal{S}_{\mathcal{M}}$.

Channel Count Saturation

Theorem 3.15. *Let $\mathcal{M}^I$ be a multichannel model such that $|I| = \dim(\mathcal{M})$. Then for a generic $S \in \mathcal{M}^I$ one has $\mathcal{S}_{\mathcal{M}^I}(S) = \mathcal{S}_{\mathcal{M}}$, that is, the weak preserving set of a generic (multichannel) S is the strong preserving set of the single-channel model $\mathcal{M}$.*

The theorem above has the corollary:

Corollary 3.16. *If a multichannel model $\mathcal{M}^L$ with $L = \dim(\mathcal{M})$ is not identifiable, then it is not identifiable for any other value of L. Such a model will be referred to as* never-identifiable. *Conversely, a single-channel model $\mathcal{M}$ with trivial strong preserving $\mathcal{S}_{\mathcal{M}}$ is* always *identifiable in a multichannel setup having $L = \dim(\mathcal{M})$ channels.*

Example 3.17. *Assume that we wish to identify a stack of L 2D discrete FIRs of size 8×8 followed by sampling* `P` *(fixed and operating concurrently on all channels). If the system is not identifiable with 64 channels, then it is never identifiable (even if we increase the number of channels). The lack of identifiability is originated in* `P`.

3.4 Mosaic Families

A *mosaic structure* is a special case of linear families spanned by disjoint indicator matrices (see Figure 3.5), for example, Toeplitz and Hankel matrices are mosaic (sometimes known as *structured matrix*). This model is particularly useful for discrete filter identifiability analysis.

We are interested in characterizing the weak and strong preserving sets of mosaic families. For the sake of the discussion, assume that $U = \mathbf{k}^{\mathcal{U}}$ and $V = \mathbf{k}^{\mathcal{V}}$ for two finite set $\mathcal{U}$ and $\mathcal{V}$, so that every operator is characterized by a matrix in $\mathbf{k}^{\mathcal{V}\times\mathcal{U}}$.

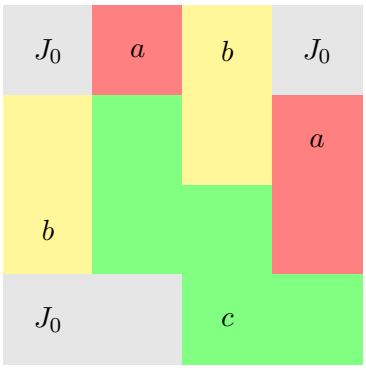


Figure 3.5: Mosaic structure with three tiles

Indicator functions take only the values 0 and 1, usually indicating membership of an element to a set. We say that a matrix S is an *indicator* if all its nonzero entries have the same value, i.e., $S_{v,u} \in \{0, \alpha\}$. Such a matrix can be thought of as an indicator function of a subset of $\mathcal{V} \times \mathcal{U}$.

Definition 3.18. *A* mosaic *on $\mathcal{V} \times \mathcal{U}$ consists of a collection $\{J_\sigma\}_{\sigma\in\mathbf{\Sigma}}$ of* disjoint *index subsets (called* tiles*) of $\mathcal{V} \times \mathcal{U}$. We define the* zero tile *of a mosaic as the collection of pairs in $\mathcal{V} \times \mathcal{U}$ uncovered by the rest of the tiles:*

$$J_0 := \mathcal{V} \times \mathcal{U} \setminus \bigsqcup_{\sigma\in\mathbf{\Sigma}} J_\sigma$$

For example, the mosaic in Figure 3.5 is defined on $|\mathcal{V}\times\mathcal{U}| = 4\times4$ and has three tiles $\mathbf{\Sigma} = \{a, b, c\}$. The sets $\{J_\sigma\}$ do not necessarily partition $\mathcal{V} \times \mathcal{U}$, unless J_0 is included as a class:

Definition 3.19. *The union $\{J_0\}\cup\{J_\sigma\}_{\sigma\in\mathbf{\Sigma}}$ is a partition of $\mathcal{V}\times\mathcal{U}$. We use the notation $\sim$ to denote equivalence under this partition, namely, $(v, u) \sim (v', u')$ iff $(v, u), (v', u') \in J_\sigma$ for some $\sigma \in \mathbf{\Sigma}$ or $(v, u), (v', u') \in \bar{J}$.*

Definition 3.20. *An operator $S \in \mathtt{Lin}(U,V)$ is* structured *with respect to a mosaic $\{J_\sigma\}_{\sigma\in\mathbf{\Sigma}}$ if the matrix values $S_{v,u} := \langle v|S|u\rangle$ are constant when $(v,u) \in J_\sigma$ and vanish on J_0, namely:*

$$S_{v,u} = S_{v',u'} \qquad (v,u) \sim (v',u') \tag{3.11}$$

$$S_{v,u} = 0 \qquad (v,u) \in J_0 \tag{3.12}$$

Note that whenever J_0 is empty, then the condition (3.12) is void. Also, if J_σ is a singleton, then the condition (3.11) is redundant for that J_σ. Another way to interpret mosaic structures is the set of all functions whose level sets are prescribed.

Mosaic operators constitute a linear subspace $\mathcal{M} \subset \mathtt{Lin}(U,V)$, which is explicitly spanned by the operators $\{S^{(\sigma)}\}_{\sigma\in\mathbf{\Sigma}}$ that are indicators of $\{J_\sigma\}_{\sigma\in\mathbf{\Sigma}}$:

$$S^{(\sigma)} = \sum_{(v,u)\in J_\sigma} |v\rangle\langle u| \tag{3.13}$$

$$S = \sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma S^{(\sigma)} = \sum_{\sigma\in\mathbf{\Sigma}} \sum_{(v,u)\in J_\sigma} \theta_\sigma |v\rangle\langle u| \tag{3.14}$$

3.4.1 $\mathcal{M}$-Preserving Pre-Compositions of Mosaic Families

We are interested in the constraints on $\mathbf{A} \in \mathtt{Lin}(U)$ that preserve a mosaic $\mathcal{M}$ (in both weak and strong senses). Plug the co-vector decomposition of $\mathbf{A}$, and the explicit sum of S in (3.14) and get

$$\begin{aligned} S\mathbf{A} &= \overbrace{\sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma \sum_{(v,u)\in J_\sigma} |v\rangle\langle u|}^{S} \overbrace{\Big(\sum_{u'\in\mathcal{U}} |u'\rangle\langle u'|\,\mathbf{A}\Big)}^{\mathbf{A}} \\ &= \sum_{\sigma\in\mathbf{\Sigma}} \sum_{(v,u)\in J_\sigma} \theta_\sigma |v\rangle \sum_{u'\in\mathcal{U}} \underbrace{\langle u|u'\rangle}_{\delta_{u,u'}} \langle u'|\,\mathbf{A} \\ &= \sum_{\sigma\in\mathbf{\Sigma}} \sum_{(v,u)\in J_\sigma} \theta_\sigma |v\rangle\langle u|\,\mathbf{A} \end{aligned} \tag{3.15}$$

Evaluating $S\mathbf{A}$ specifically at entry (v',u') yields

$$(SA)_{v',u'} = \sum_{\sigma\in\mathbf{\Sigma}} \sum_{(v,u)\in J_\sigma} \theta_\sigma \delta_{v',v} \langle u|\mathbf{A}|u'\rangle = \sum_{\sigma\in\mathbf{\Sigma}} \sum_{(v',u)\in J_\sigma} \theta_\sigma \mathbf{A}_{u,u'} \tag{3.16}$$

Plug (3.16) into (3.11) and (3.12) to determine constraints on $\mathbf{A}$:

Weak Preserving Variety of Mosaic Structure

The pre-composition $\mathbf{A}$ is preserving a specific $\mathbf{m}(\theta)$ in $\mathcal{M}$ whenever:

$$\sum_{\sigma\in\mathbf{\Sigma}}\theta_\sigma\left[\sum_{(v',u'')\in J_\sigma}\mathbf{A}_{u'',u'}-\sum_{(v,u'')\in J_\sigma}\mathbf{A}_{u'',u}\right]=0\quad(v,u)\sim(v',u')$$

$$\sum_{\sigma\in\mathbf{\Sigma}}\sum_{(v,u')\in J_\sigma}\theta_\sigma\mathbf{A}_{u',u}=0\qquad(v,u)\in J_0$$

and the strong condition

Strong Preserving Variety of Mosaic Structure

The pre-composition $\mathbf{A}$ is preserving $\mathcal{M}$ whenever for every $\sigma\in\mathbf{\Sigma}$:

$$\sum_{(v',u'')\in J_\sigma}\mathbf{A}_{u'',u'}=\sum_{(v,u'')\in J_\sigma}\mathbf{A}_{u'',u}\qquad(v,u)\sim(v',u'),$$

$$\sum_{(v,u')\in J_\sigma}\mathbf{A}_{u',u}=0\qquad(v,u)\in J_0$$

In Chapter 4, we will further characterize the strong and weak preserving sets of mosaic families corresponding to FIR convolution operators, and provide some examples for the weak and strong $\mathcal{M}$-preserving equations. This characterization is useful to study the identifiability (or lack thereof) of uniformly down-sampled discrete systems.

CHAPTER 4

DISCRETE FIR FILTERS

This chapter is devoted to study identifiability of discrete multichannel finite impulse response (FIR) convolutions subject to sampling, of the type depicted in Example 1.1. This will be done both in the parameter space, as well as $\mathcal{M}$-preserving pre-composition. The identification guarantee will be utilized to test identifiability of different sampling configurations. Then, we will show that sub-sampling on index subgroups is never identifiable.

4.1 The Discrete FIR Model

We consider signals of the type $x \in \mathbf{k}^{\mathbb{T}}$, here $\mathbb{T}$ is a discrete (finitely generated) Abelian group (for example: $\mathbb{Z}^N$ or $\mathbb{Z}_N$) to the field $\mathbf{k}$. The signal value at index $t \in \mathbb{T}$ will be denoted either $x[t]$ or $\langle t|x\rangle$. The space $\mathbf{k}^{\mathbb{T}}$ is not necessarily finite-dimensional (depending on the cardinality of the group $\mathbb{T}$, e.g. $\mathbb{Z}$ or $\mathbb{Z}^2$ often used in DSP are infinite). For an index subset $I \subset \mathbb{T}$, we let $\langle I\rangle \subset \mathbf{k}^{\mathbb{T}}$ denote the subspace of all finitely supported functions with support limited to I.

For any $\tau \in \mathbb{T}$, the *shift operator* $\varsigma^{\mathbb{T}} \in \mathtt{Lin}(\mathbf{k}^{\mathbb{T}})$ if defined by

$$(\varsigma^{\tau} x)[t] := x[t+\tau] \tag{4.1}$$

and can be written also as $\varsigma^{\tau} = \sum_{t\in\mathbb{T}} |t\rangle \langle t+\tau|$. All shift operators commute and have inverses, and form a group isomorphic to $\mathbb{T}$ (by $\tau \mapsto \varsigma^{\tau}$). Shifting extends naturally to multichannel setup $\mathbf{k}^L \otimes \mathbf{k}^{\mathbb{T}}$ by acting on all channels simultaneously.

Every finitely supported $\mathbf{h} \in \ell_0(\mathbf{k}^{\mathbb{T}})$ has a corresponding *convolution operator* $\mathtt{C}_{\mathbf{h}} \in \mathtt{Lin}(\mathbf{k}^{\mathbb{T}})$ defined by the (finite) sum

$$\mathtt{C}_{\mathbf{h}} := \sum_{\tau\in\mathbb{T}} \mathbf{h}[\tau]\varsigma^{-\tau} \tag{4.2}$$

This definition extends seamlessly to multichannel FIR convolution, where $\mathbf{h} \in \mathbf{k}^L \otimes \ell_0(\mathbf{k}^{\mathbb{T}})$, but then the operator maps between $\mathtt{C}_{\mathbf{h}} : \mathbf{k}^{\mathbb{T}} \to \mathbf{k}^L \otimes \mathbf{k}^{\mathbb{T}}$.

For an index subset $I \subset \mathbb{T}$, the *sampling operator* $\mathtt{D}_I : \mathbf{k}^{\mathbb{T}} \to \mathbf{k}^I$ is merely the restriction on I:

$$\mathtt{D}_I x := x\big|_I \tag{4.3}$$

which is linear. Continuing along that line, we define the *sampling projection* $\mathtt{P}_I \in \mathtt{Lin}(\mathbf{k}^{\mathbb{T}})$ as projection:

$$(\mathtt{P}_I x)[t] := \begin{cases} x[t] & t \in I \\ 0 & \text{else} \end{cases} \tag{4.4}$$

When $\mathbf{k}^{\mathbb{T}}$ has an inner product, it is easy to show that the conjugate $\mathtt{D}_I : \mathbf{k}^I \to \mathbf{k}^{\mathbb{T}}$ is the zero-padding operator

$$(\mathtt{D}_I^* y)[t] := \begin{cases} y[t] & t \in I \\ 0 & \text{else} \end{cases} \tag{4.5}$$

and that in general $\mathtt{P}_I = \mathtt{D}_I \mathtt{D}_I^*$.

For a multichannel tensor $y \in \mathbf{k}^L \otimes \mathbf{k}^{\mathbb{T}}$, the restriction/projection act on all L channels simultaneously (i.e. $\mathtt{D}_I$ maps into $\mathbf{k}^L \otimes \mathbf{k}^I$).

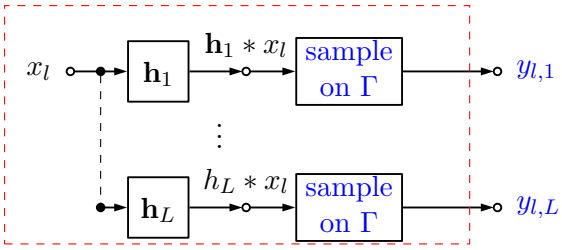


Figure 4.1: Multichannel convolution with sampling

Assume that a finite set of signals $\{y_i\}_{i \in \mathcal{E}}$ are outputs of samples multichannel FIR system (see Figure 4.1), that is,

$$y_i = \mathtt{D}_\Gamma \mathtt{C}_{\mathbf{h}} x_i, \quad i \in \mathcal{E} \tag{4.6}$$

Here the subset $\Gamma \subset \mathbb{T}$ (the *sampling pattern*) is finite, $\mathbf{h} \in \mathbf{k}^L \otimes \mathbf{k}^{\mathbb{T}}$ is an unknown FIR sequence, and $\{x_i\}_{i \in \mathcal{E}}$ are some *unknown* input signals.

The recovery of $\mathbf{h}$ from the samples in (4.6) is known as the *blind multichannel FIR identification problem*, and the joint recovery of h and x_α is known as *blind FIR deconvolution*.

Problem Statement

What filter values $\mathbf{h}$ and sampling patterns Γ can be identified by output spans?

In this chapter we essentially answer this question by providing various conditions (necessary or sufficient).

4.2 Restricted FIR Support Configurations

This problem can in fact be formulated as a special case of range-space based system identification of the parametric model $\mathbf{h} \mapsto \mathtt{D}_\Gamma \mathtt{C}_\mathbf{h}$. Nevertheless, there are two obstacles:

1. The dimension $\dim(\mathbf{k}^\mathbb{T}) = |\mathbb{T}|$ is not necessarily finite.
2. There is an inherent shift ambiguity: for every $\tau \in \mathbb{T}$ one has $\mathtt{C}_{\varsigma^\tau \mathbf{h}} = \mathtt{C}_\mathbf{h} \varsigma^\tau$, thus $\mathtt{Im}(\mathtt{C}_\mathbf{h}) = \mathtt{Im}(\mathtt{C}_{\varsigma^\tau \mathbf{h}})$ while usually $\mathbf{h} \neq \varsigma^\tau \mathbf{h}$.

Both issues can be addressed by restricting the support of $\mathbf{h}$ to a specified *finite domain*

$$\Sigma \subset \mathbb{T} \qquad \text{(FIR Domain Restriction)}$$

namely, $\mathbf{h} \in \mathbf{k}^L \otimes \langle \Sigma \rangle$. The space $\mathbf{k}^L \otimes \langle \Sigma \rangle \cong \mathbf{k}^L \otimes \mathbf{k}^\Sigma$ is finite-dimensional, and furthermore, the output $\mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} x$ depends on the input x only on the *dilation* (*Minkowsky difference*)

$$\Gamma - \Sigma := \{\gamma - \sigma \;\; | \;\; \sigma \in \Sigma, \;\; \gamma \in \Gamma\} \tag{4.7}$$

so that

$$\mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} = \mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} \mathtt{P}_{\Gamma - \Sigma} \tag{4.8}$$

is a finite rank operator with input dimension $|\Gamma - \Sigma|$ and output dimension $L \times |\Gamma|$, and fits the apparatus we have.

Our next step is to define a parametric family. Fix the parameter space $\mathbf{M} = \mathbf{k}^L \otimes \mathbf{k}^\Sigma$. Every parameter $\theta \in \mathbf{k}^L \otimes \mathbf{k}^\Sigma$ corresponds to a unique FIR $\mathbf{h}(\theta) \in \mathbf{k}^L \otimes \langle \Sigma \rangle$ (denoted just $\mathbf{h}$) by the bijection

$$\mathbf{h}[\sigma] = \theta[\sigma] \quad \text{for all } \sigma \in \Sigma$$

The difference between $\mathbf{h}$ and θ is technical: the former is also defined outside Σ, and vanish there. Then, define the parametric family as follows:

$$\begin{aligned} \mathbf{m}(\theta) &: \underbrace{\langle \Gamma - \Sigma \rangle}_{U} \to \underbrace{\mathbf{k}^L \otimes \mathbf{k}^\Gamma}_{V} \\ \mathbf{m}(\theta) &:= \mathtt{D}_\Gamma \mathtt{C}_{\mathbf{h}(\theta)} \end{aligned} \tag{4.9}$$

Equation (4.6) now becomes $y_i = \mathbf{m}(\theta)x_i$, and our goal is to identify θ.

Definition 4.1. *We call the triplet* (Σ, Γ, L) *a* an FIR configuration *(or just configuration), with a corresponding algebraic family* (4.9)*.*

4.2.1 Identifiability of Sampled FIR Configurations

Theorem 2.24 suggest that identifiability is a generic property (at least in the complex case): if there is one identifiable FIR filter $\mathbf{h}$, then almost all of them are. Therefore, the term *identifiable configuration* is meaningful (in the sense that the associated algebraic family is almost everywhere identifiable).

The choice of the configuration parameters affects identifiably as follows:

- **The choice of FIR domain Σ:** When Σ is too small then $\mathbf{m}$ is not surjective on V, thus a solution $\mathbf{h}$ may not exist for a given subspace in $\mathtt{Gr}(r, V)$. For example, choosing $\Sigma = \{0\}$ imposes $\mathbf{h}$ to be an impulse:

  $$y_i = \mathbf{h}[0]\mathtt{D}_\Gamma x_i$$

  This holds true only if $\{y_i\}$ are all co-linear. On the flip side, if Σ is large enough so that $\Sigma \supset \Gamma$, then a solution always exists (pick $\mathbf{h}[t] = y[t]$ and $x[t] = \delta_0[t]$, a unit impulse at $t = 0$) but is not necessarily unique, as Σ may support different shifts of $\mathbf{h}$.

- **The choice of output sampling pattern Γ:** Identifiability is affect by the size or structure of Γ. For $\mathbb{T} = \mathbb{Z}$ (i.e. sequences) where Σ and Γ are integer intervals, identifiability is determined by their lengths (see [2] and [4]). In general, if Γ is decimated (subsample at a uniform rate) then the identification always exhibits certain ambiguity (see [7] for $\mathbb{T} = \mathbb{Z}^2$, and Theorem 4.10 for more general index groups). Our work extends the existing literature by allowing general (and unstructured) sets Γ and Σ on general index groups, where studying identifiability in terms of the structure of (Σ, Γ) is quite difficult.

- **The number of channels** L: Generally, more channels mean less ambiguity, up to the limit of $L = |\Sigma|$.

4.3 Searching for Identifiable Configurations

One of the basic problems motivating this research was to search for identifiable configurations.

Testing for identifiability can be done using numerical guarantee derived in Theorem 2.24 for a randomly selected parameter. An exhaustive search of all patterns partial to a finite set $\bar{\Gamma} \subset \mathbb{T}$ has exponential complexity, and amounts to testing $2^{|\bar{\Gamma}|}$ configurations (per *fixed* selection of Σ and L). For example, an exhaustive search for identifiable configurations limited to 32×32 pixels in $\mathbb{Z}^2$ has 2^{1024} iterations, and is clearly impractical.

The good news is that testing every pattern is not necessary, nonetheless, and many patterns can be sifted out by necessary conditions; for example we can require

$$|\Gamma - \Sigma| < L \cdot |\Gamma| \tag{4.10}$$

for otherwise the algebraic condition of "tall" matrices is not satisfied. Also, if $\Gamma \subset \tilde{\Gamma}$ for a non-identifiable configuration $(\Sigma, \tilde{\Gamma}, L)$, less data is not going to help (see Theorem 4.2).

Some manipulations of configurations maintain parameter identifiability, as described in the following result (see proof in Appendix A).

Theorem 4.2. *Let (Σ, Γ, L) be an identifiable configuration. Then the following configurations are identifiable:*

1. Oversampling: $(\Sigma, \tilde{\Gamma}, L)$ *where* $\tilde{\Gamma} \supset \Gamma$.
2. Shifting: $(\tau + \Sigma, \gamma + \Gamma, L)$ *for every* $\tau, \gamma \in \mathbb{T}$.
3. Index Automorphisms*:* $(\Phi\Sigma, \Phi\Gamma, L)$ *where* $\Phi : \mathbb{T} \to \mathbb{T}$ *is any group automorphism.*

See proof in Appendix A.

Thus, once a pattern Γ is pronounced identifiable, then all patterns containing Γ are also identifiable. Likewise, all shifts of Γ are identifiable, and all reflections/symmetries of Γ remain identifiable (for a given Σ, L).

Another way to rule out a configuration (Σ, Γ, L) is testing whether its *single-channel* counterpart (Σ, Γ, L) is never identifiable, essentially by computing the corresponding strong preserving family $\mathcal{S}_{\mathcal{M}}$.

4.4 $\mathcal{S}_{\mathcal{M}}$ of Discrete FIR Families

We return to study $\mathcal{S}_{\mathcal{M}}$ corresponding to a FIR channel associated with the configuration (Σ, Γ, L). Recall that if the strong preserving set $\mathcal{S}_{\mathcal{M}} \neq \text{span}\{\texttt{Id}\}$ of the single-channel model, then no multichannel extension will be identifiable. Having said that, we shall have a look at the structure of $\mathcal{S}_{\mathcal{M}}$ for a single-channel configuration $(\Sigma, \Gamma, 1)$

Let $\mathcal{M}$ be the (single-channel) FIR model associated with (Σ, Γ). Recall that $\mathbf{A} \in \mathcal{S}_{\mathcal{M}}$ merely amounts to column operation on $\mathbf{m}(\theta)$. A simple case of $\mathbf{A}$ is just a column move from column $\tau_1 \in \Gamma - \Sigma$ to another column $\tau_2 \in \Gamma - \Sigma$. However, there is no guarantee that such simple moves preserve a sampled FIR structure. Usually, preserving $\mathbf{A}$ involves a complex linear shuffle of the columns of the operator $\mathbf{m}(\theta)$. We will try to find a simple basis for $\mathcal{S}_{\mathcal{M}}$, comprised of preserving pre-compositions that are as close as possible to column shuffle. This can be done thanks to the special Toeplitz structures of FIR operators. Every FIR structure (generalized Toelitz), whether sampled or not, is a mosaic.

Definition 4.3. *A mosaic* $\{J_\sigma\}$ *is called* FIR mosaic *if it corresponds to a convolution operator.*

All involved index sets $\mathbf{\Sigma}, \mathcal{U}, \mathcal{V}$ of FIR-mosaic are subsets $\mathbb{T}$ (not necessarily subgroups/cosets), where $\mathcal{U} = \mathcal{V} - \mathbf{\Sigma}$ (a Minkowski difference) and

$$J_\sigma = \{(\gamma, v - \sigma) \quad | \quad v \in \mathcal{V}\} \tag{4.11}$$

A slight relaxation of an FIR mosaic is an *admissible mosaic*:

Definition 4.4. *A mosaic* $\{J_\sigma\}$ *is* admissible mosaic *if:*

1. *Every tile* $J_\sigma \subset \mathcal{V} \times \mathcal{U}$ *is a* graph of a one-to-one function in v*:*

$$J_\sigma = \{(v, u_\sigma(v)) \mid v \in \mathcal{V}\}$$

   Here $u_\sigma : \mathcal{V} \to \mathcal{U}$ *is injective. Every* v *has exactly one paired* u *in* J_σ

2. *Every* $u \in \mathcal{U}$ *belongs to at least one tile* J_σ*.*

An admissible mosaic is fully characterized by the functions $\{u_\sigma \mid \sigma \in \mathbf{\Sigma}\}$.

The mosaic in Figure 4.2 is admissible. Each color represents a different class in $\{J_\sigma\}$, and the white cells are J_0. The mosaic in Figure 3.5 is not admissible.

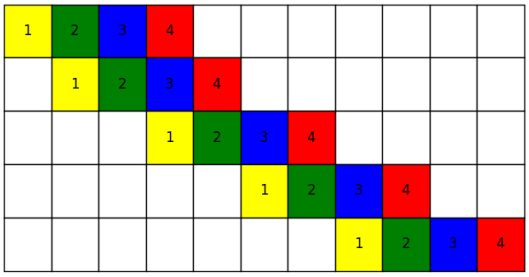


Figure 4.2: Admissible mosaic

We can easily show that FIR mosaics are also admissible (see proof in Appendix A):

Proposition 4.5. *An FIR mosaic is admissible.*

The preserving pre-compositions of admissible mosaic families are fairly simple to characterize. Equation (3.16) reduces to $(S\mathbf{A})_{v,u} = \sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma \mathbf{A}_{u_\sigma(v),u}$ so the weak structure condition now becomes:

$$\sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma \left[\mathbf{A}_{u_\sigma(v'),u'} - \mathbf{A}_{u_\sigma(v),u}\right] = 0 \qquad (v,u) \sim (v',u')$$

$$\sum_{\sigma\in\mathbf{\Sigma}} \theta_\sigma A_{u_\sigma(v),u} = 0 \qquad (v,u) \in J_0$$

Furthermore, note that the condition $(v,u) \sim (v',u')$, can be replaced with $u = u_\sigma(v)$ and $u' = u_\sigma(v')$ for some $\sigma \in \mathbf{\Sigma}$, so in conclusion:

Weak Preserving Condition of Admissible Mosaic

For admissible mosaic family $\mathcal{M}$, the weak $\mathcal{S}_\mathcal{M}(S)$ (a linear variety) is characterized by all $\mathbf{A}$ satisfying

$$\sum_{\sigma'\in\mathbf{\Sigma}} \theta_{\sigma'} \left[\mathbf{A}_{u_{\sigma'}(v'),u_{\sigma'}(v')} - \mathbf{A}_{u_\sigma(v),u_\sigma(v)}\right] = 0 \quad v,v' \in \mathcal{V},\ \sigma \in \mathbf{\Sigma} \quad (4.12)$$

$$\sum_{\sigma'\in\mathbf{\Sigma}} \theta_{\sigma'} \mathbf{A}_{u_{\sigma'}(v),u} = 0 \qquad (v,u) \in J_0 \quad (4.13)$$

For the strong mosaic-preserving, we require that the latter holds for every basis vector $S = S_\sigma$, and the equation becomes

Strong Preserving Condition of Admissible Mosaic

The strong preserving $\mathcal{S}_\mathcal{M}$ of an admissible mosaic $\mathcal{M}$ are all $\mathbf{A}$ s.t:

$$\mathbf{A}_{u_{\sigma'}(v'),u_\sigma(v')} = \mathbf{A}_{u_{\sigma'}(v),u_\sigma(v)} \qquad v, v' \in \mathcal{V}, \quad \sigma', \sigma \in \mathbf{\Sigma} \tag{4.14}$$

$$\mathbf{A}_{u_\sigma(v),u} = 0 \qquad (v,u) \in J_0, \quad \sigma \in \mathbf{\Sigma} \tag{4.15}$$

The latter condition is in fact a mosaic condition:

Lemma 4.6. *For an admissible mosaic structure $\mathcal{M}$, the set $\mathcal{S}_\mathcal{M}$ of strong preserving pre-compositions is itself a mosaic structure $\{\tilde{J}_k\}$ on $\mathcal{U} \times \mathcal{U}$. We shall refer to $\{\tilde{J}_k\}$ as the* pre-composition mosaic.

See Appendix A for proof. Figure 4.3 demonstrates a FIR mosaic family and its corresponding strong pre-composition mosaic.

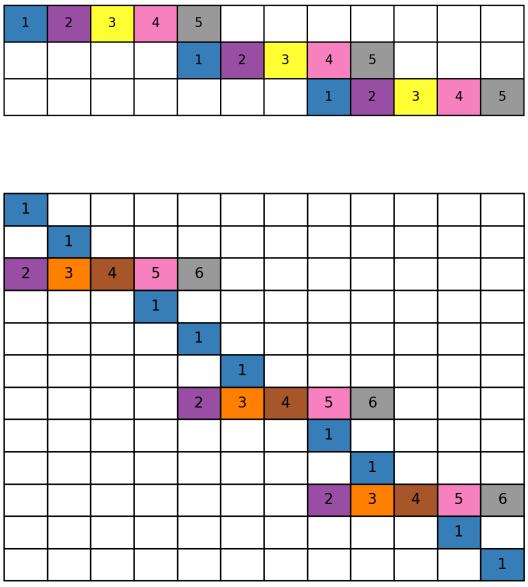


Figure 4.3: FIR mosaic (top) and its $\mathcal{S}_\mathcal{M}$ pre-composition mosaic (bottom)

We turn to study the tiles $\{\tilde{J}_k\}$, whose corresponding mosaic operators $\mathbf{A}^{(k)} \in \texttt{Lin}(U)$ defined by $\mathbf{A}^{(k)} := \sum_{(u,u')\in\tilde{J}_k} |u\rangle \langle u'|$ span $\mathcal{S}_\mathcal{M}$. One can think of $\mathbf{A}^{(k)}$ as an adjacency matrix of a directed graph $\mathcal{G}_k := (\mathcal{U}, J_k)$, where the pair $(u,u') \in \mathcal{U}^2$ is connected by $u' \mapsto u$, whenever $(u,u') \in J_k$.

The implication of Lemma 4.6 is that for an admissible mosaic family $\mathcal{M}$, the space $\mathcal{S}_\mathcal{M}$ is spanned by indicator matrices. Any indicator pre-composition $\mathbf{A}$ corresponds to a directed graph on the columns $\mathcal{U}$.

Finally, for FIR mosaic, we have the following:

Theorem 4.7. *For an FIR mosaic $\{J_\sigma\}$, every nonzero tile of the pre-composition mosaic $\tilde{J}_k$ (i.e. whose elements do not admit the zero constraint) has at most one class representative per entry, that is,*

$$(u, u'), (u, u'') \in \tilde{J}_k \quad \Rightarrow \quad u' = u'' \tag{4.16}$$

$$\text{and} \quad (u', u), (u'', u) \in \tilde{J}_k \quad \Rightarrow \quad u' = u'' \tag{4.17}$$

See Appendix A for a proof.

A Basis for Strong FIR-Preserving Pre-compositions

Corollary 4.8. *The space $\mathcal{S}_\mathcal{M}$ of (strong) preserving pre-composition of an FIR model $\mathcal{M}$ is spanned by graphs whose nodes have degree 0 or 2, any edge is connected to exactly two nodes (allowing self-edges $u \to u$). Thus, to check whether some FIR model is identifiable, it is sufficient to find which column moves of its matrix do not alter its structure (hence keeping it in $\mathcal{M}$).*

Example 4.9. *We demonstrate the Corollary 4.8 on the configuration $\Sigma = \{0, \ldots, 5\}$, $\Gamma = \{0, 3, 8\}$, and $L = 6$. Empirical testing of a random S_0 shows that $\dim(\mathcal{S}_\mathcal{M}(S_0)) = 4$, with four basis matrices (see Figure 4.4). We shall verify that result by the tedious combinatorial task of finding which column shifts preserve the matrix structure. This process is similar to Example 3.6, except that now we know that column shifts constitute a basis of $\mathcal{S}_\mathcal{M}$.*

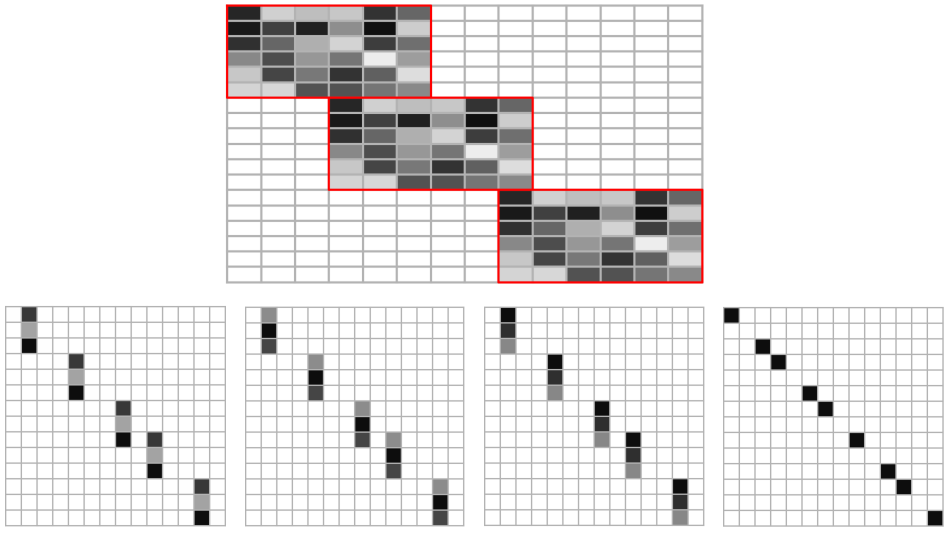


Figure 4.4: Top: Matrix structure of the family in Example 4.9. Bottom: Empirically computed basis for $\mathcal{S}_\mathcal{M}(S_0)$

A column shift, denoted $\sigma \to \sigma'$ where $\sigma, \sigma' \in \Sigma$, is an edge in the directed graph $\Sigma \times \Sigma$. There are $|\Sigma|^2 = 36$ possible moves in this example.[1] *Any column shift may entail other shifts, based on the connectivity graph in Figure 4.5. Entailed moves that land outside Σ (denoted $\{\sigma \to nil\}$) are deemed* invalid. *Moves can be grouped into equivalence classes by entailment relation, where a class is valid if all its moves are valid (see Table 4.1).*

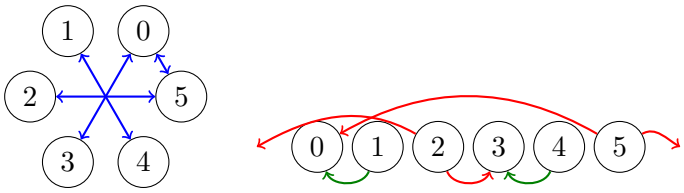


Figure 4.5: Connectivity graph of Σ in Example 4.9 (left), column shifts and their entailed moves (right). Allowed moves in green, otherwise in red

Table 4.1: Attempted moves and entailed moves

Attempted Move	**Entailed Moves**	**Validity**
$\{0 \to 1, \ldots, 5\}$	$\{5 \to \text{nil}\}$	not allowed
$\{1 \to 0, 2\}$	$\{4 \to 3, 5\}$	allowed
$\{1 \to 3, 4, 5\}$	$\{4 \to \text{nil}\}$	not allowed
$\{2 \to 0, 1\}$	$\{5 \to 3, 4\} \Rightarrow \{0 \to \text{nil}\}$	not allowed
$\{2 \to 3, 4, 5\}$	$\{5 \to \text{nil}\}$	not allowed
$\{3 \to 0, 1, 2\}$	$\{0 \to \text{nil}\}$	not allowed
$\{3 \to 4, 5\}$	$\{0 \to 1, 2\} \Rightarrow \{5 \to \text{nil}\}$	not allowed
$\{4 \to 0, 1, 2\}$	$\{1 \to \text{nil}\}$	not allowed
$\{4 \to 3, 5\}$	$\{1 \to 0, 2\}$	allowed
$\{5 \to 0, \ldots, 4\}$	$\{0 \to \text{nil}\}$	not allowed
$\{0 \to 0, 2 \to 2, 3 \to 3, 5 \to 5\}$	$\{0 \to 0, 2 \to 2, 3 \to 3, 5 \to 5\}$	allowed
$\{1 \to 1, 4 \to 4\}$	$\{1 \to 1, 4 \to 4\}$	allowed

There are four valid classes in this example, two column shuffle classes $\{1 \to 0,\ 4 \to 3\}$ and $\{4 \to 5,\ 1 \to 2\}$, and two loop classes $\{1 \to 1,\ 4 \to 4\}$ and $\{0 \to 0,\ 2 \to 2,\ 3 \to 3,\ 5 \to 5\}$. Those four classes correspond to four pre-compositions in $\mathcal{S}_{\mathcal{M}}$, which constitute a basis for $\mathcal{S}_{\mathcal{M}}$ due to Corollary 4.8, hence $\dim(\mathcal{S}_{\mathcal{M}}) = 4$, conforming to the empirical test.

[1] For larger models, this process has been automated by a script written in Python.

4.4.1 $\mathcal{S}_{\mathcal{M}}$ of Uniform Subsampling

Finally, we show that FIR systems followed by uniform sub-sampling are never identifiable (algebraically).

Theorem 4.10. *Let $G \subset \mathbb{T}$ be a proper subgroup. A configuration (Σ, Γ, L) with non-singleton Σ and $\Gamma \subset \gamma_0 + G$ (a coset) is never identifiable (regardless of L). Furthermore, if the index of the subgroup (i.e. "sampling rate") satisfies $1 < [\mathbb{T} : G] \leq |\Sigma|$, then $\dim(\mathcal{S}_{\mathcal{M}}) \geq [\mathbb{T} : G]^2$.*

The Case Against Uniform Sampling

The practical implication is that uniform sampling always produces filter ambiguity. For example, sampling $\mathbb{Z}$ at rate d has quotient group $\mathbb{Z}/d\mathbb{Z} = \mathbb{Z}_d$ resulting with $\dim(\mathcal{S}_{\mathcal{M}}) \geq d^2$. Sampling $\mathbb{Z}^2$ at rate d makes $\mathbb{Z}^2/(d\mathbb{Z} \oplus d\mathbb{Z}) = \mathbb{Z}_d \oplus \mathbb{Z}_d$ so that the dimension of the ambiguity is at least $\dim(\mathcal{S}_{\mathcal{M}}) \geq (d \cdot d)^2 = d^4$.

Proof. (Theorem 4.10) First note that (Σ, Γ, L) is identifiable if and only if $(\Sigma, \Gamma - \gamma_0, L)$ is identifiable, so we can assume without loss of generality that Γ is a subset of G, rather than the coset $\gamma_0 + G$. The group $\mathbb{T}$ with a subgroup G admits the decomposition

$$\mathbb{T} \cong P \oplus G$$

where the quotient $P := \mathbb{T}/G$ is itself an Abelian group. Every $t \in \mathbb{T}$ admits a unique decomposition in $P \oplus G$ given by

$$t = r + p, \qquad p \in P, \quad r \in G \tag{4.18}$$

Here p ("phase") corresponds to the coset $p+G \in P$, and can be represented by any arbitrary element of the $p+G$. This group decomposition gives rise to the so-called polyphase decompositions $s_p, \bar{s}_p \in \mathbf{k}^G$ defined for every signal $s \in \mathbf{k}^{\mathbb{T}}$ by the relations

$$s_p[r] := s[r - p] \qquad \text{(type I)} \tag{4.19}$$

$$\bar{s}_p[r] := s[r + p] \qquad \text{(type II)} \tag{4.20}$$

for every $r \in G$ and $p \in P$. The signal $s[t]$ can be then expressed as:

$$s[t] = \sum_{p \in P} s_p[t + p] = \sum_{p \in P} \bar{s}_p[t - p]$$

Note that if $s[t]$ is finitely supported, then the polyphase components are finitely supported as well, and vice versa. For $\tau \in G$, substitute $t = r + p$ per (4.18) in the convolution sum, then expand to a double sum over P and G, and plug in (4.19) and (4.20):

$$\begin{aligned}(\mathtt{C}_{\mathbf{h}}x)[\tau] &= \sum_{t\in\mathbb{T}} \mathbf{h}[\tau - t]x[t] = \sum_{p\in P}\sum_{r\in G} \underbrace{\mathbf{h}[\overbrace{\tau - r}^{\in G} - p]}_{\mathbf{h}_p[\tau - r]} \underbrace{x[r + p]}_{\bar{x}_p[r]} \\ &= \sum_{p\in P} (\mathbf{h}_p * \bar{x}_p)[\tau] = \sum_{p\in P} (\mathtt{C}_{\mathbf{h}_p} x_p)[\tau]\end{aligned}$$

The convolution in the second line is done over the (discrete, Abelian) subgroup G, i.e. $\mathtt{C}_{\mathbf{h}_p} \in \mathtt{Lin}(\mathbf{k}^{\mathbb{T}})$. For a nonsingular $B \in \mathbf{k}^{P\times P}$ and $s \in \mathbf{k}^{\mathbb{T}}$, define the polyphase *shuffles* $\psi_B(s)$ and $\bar{\psi}_B(s)$ as

$$(\psi_B s)_{p'} := \sum_{p\in P} B_{p,p'} s_p \tag{4.21}$$

$$(\bar{\psi}_B s)_{p'} := \sum_{p\in P} B_{p,p'} \bar{s}_p \tag{4.22}$$

The mappings ψ_B and $\bar{\psi}_B$ are linear and keep their argument support ($\psi_B \langle\Sigma\rangle \subset \langle\Sigma\rangle$). Let $\mathbf{g} = \psi_B(\mathbf{h})$ and $\tilde{x} = \bar{\psi}_{B^{-1}}(x)$. It is easy to verify that

$$\sum_{p\in P} \mathbf{h}_p * \bar{x}_p = \sum_{p\in P} \mathbf{g}_p * \tilde{x}_p$$

so that

$$\mathtt{D}_\Gamma \mathtt{C}_{\mathbf{h}} = \mathtt{D}_\Gamma \mathtt{C}_{\mathbf{g}} \bar{\psi}_{B^{-1}} \tag{4.23}$$

for every invertible B. In other words $\bar{\psi}_{B^{-1}} \in \mathcal{S}_{\mathcal{M}}$ is a nontrivial strong $\mathcal{M}$-preserving pre-composition, therefore such a system is never identifiable. All that is needed to lose identifiability is a single B for which $\psi_B(\mathbf{h})$ and $\mathbf{h}$ are not proportional, which always happen unless Σ is a singleton.

Whenever $N = |P| = [\mathbb{T} : G]$ is finite, we have N^2 invertible matrices in $\mathbf{k}^{P\times P}$ that are linearly independent, leading to $|N|^2$ linearly independent pre-composition ψ_B, hence $\dim(\mathcal{S}_{\mathcal{M}}) \geq |N|^2$ whenever $N \leq |\Sigma|$. Once N hits the value $|\Sigma|^2$, then the entire space $\mathtt{Lin}(U)$ makes $\mathcal{M}$-preserving pre-compositions. □

Even if $[\mathbb{T} : G]$ is infinite one can apply similar transformations on any finite subset of coordinates.

4.5 FIR Identification with Input-Output Remapping

Previously, we discussed the detection of a single FIR channel $\mathbf{h}$ by sampling small patches of neighboring data on Γ, and created input diversity by shifting it all around the index group.

The operator model $\mathbf{m}(\theta) = \mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} \mathtt{P}_{\Gamma-\Sigma}$ is usually not tall for a single-channel $\mathbf{h}$. For example, the single-channel two-tap filter $\mathbf{m} : \mathbb{C}^2 \to \mathbb{C}^{3\times 4}$

$$\mathbf{m}(\theta_0, \theta_1) = \begin{bmatrix} \theta_0 & \theta_1 & 0 & 0 \\ 0 & \theta_0 & \theta_1 & 0 \\ 0 & 0 & \theta_0 & \theta_1 \end{bmatrix} \tag{4.24}$$

is not identifiable. Increasing Γ over 3 would not help since inevitably the number of columns $|\Sigma - \Gamma|$ increases (the matrix above will forever be wide). We mitigated that by extending the problem to multiple channels, i.e. identifying several systems working concurrently on the same input.

Another way to augment a system is by introducing known pre-filters $\{S_\gamma\} \subset \mathtt{Lin}(U)$ and post-filters $\{T_\gamma\} \subset \mathtt{Lin}(V)$. The augmented system $\tilde{\mathbf{m}} : U \to \oplus_{\gamma\in\Gamma}(T_\gamma V)$ defined by

$$\tilde{\mathbf{m}}(\theta)_\gamma = T_\gamma \mathbf{m}(\theta) S_\gamma$$

can in fact be identifiable, even if $\mathbf{m}$ itself is not, as it increases the range dimension without incurred increased input dimension.

There are various ways to choose S_γ, T_γ. An interesting class of linear filters is defined by index transformations of $\mathbb{T}$. Let $\varphi : \mathbb{T} \to \mathbb{T}$ (not necessarily a group homomorphism), and define $S_\varphi x = x \circ \varphi$. We already experimented with shift $\varphi = \varsigma^\tau$, but nothing prevents us from attempting other transforms. For example, let $\varphi : \mathbb{T} \to \mathbb{T}$ be a group endomorphism onto $\mathbb{T}$, and let $\mathtt{Ker}(\varphi)$ denote its kernel. For every coset ("phase") $p \in \mathbb{T}/\mathtt{Ker}(\mathbb{T})$, define the associated *dilation* $\mathcal{L}_{\varphi,p} : V^\mathbb{T} \to V^\mathbb{T}$ (or polyphase component) as

$$(\mathcal{L}_{\varphi,p} x)[t] = x[\varphi(t) + p] \tag{4.25}$$

Example 4.11. *Let $\mathbb{T} = \mathbb{Z}$ with the group homomorphism $\phi(n) = kn$. The set of all phases is $\mathbb{Z}/(k\mathbb{Z})$ which is $\mathbb{Z}_k$, and the dilation*

$$(\mathcal{L}_{\varphi,p} x)[n] = x[kn + p]$$

is merely the p-th phase of the polyphase decomposition of x.

Intuitively, dilations amount to zooming outside an input signal, flipping it, or rotating it (if the group is rich enough).

Example 4.12. *There is a noble identity* $\mathbf{h} * \mathcal{L}_{\varphi,p}x = \mathcal{L}_{\varphi,p}(\tilde{\mathbf{h}} * x)$ *with a choice of appropriate filter* $\tilde{\mathbf{h}}_\varphi$*:*

$$\begin{aligned}(\mathbf{h} * \mathcal{L}_{\varphi,p}x)[\gamma] &= \sum_\sigma \mathbf{h}[\sigma](\mathcal{L}_{\varphi,p}x)[\gamma - \sigma] \\ &= \sum_\sigma \mathbf{h}[\sigma]x[\varphi(\gamma - \sigma) + p] \\ &= \sum_{\sigma \in \varphi^* \Sigma} \tilde{\mathbf{h}}_\varphi[\sigma]x[\varphi(\gamma) - \sigma + p] \\ &= \mathcal{L}_{\varphi,p}(\tilde{\mathbf{h}}_\varphi * x)\end{aligned}$$

Sampling $\mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} \mathcal{L}_{\varphi,p}$ *entails sampling the input* x *on the domain* $\varphi(\Gamma - \Sigma) + p$*. This materialize in practice by "spreading" the filter values differently across the matrix, creating new sampling rows.*

Consider the FIR in (4.24)*. Dilations of order* 2 *can give measurements of the form*

$$\begin{bmatrix} \theta_0 & 0 & \theta_1 & 0 \\ 0 & \theta_0 & 0 & \theta_1 \end{bmatrix}$$

(the first and second rows correspond to two different phases). Dilation of order 3 *gives the output*

$$\begin{bmatrix} \theta_0 & 0 & 0 & \theta_1 \end{bmatrix}$$

Dilations of order 4 *and above amount to scaling, i.e.* $\begin{bmatrix} \theta_0 & 0 & 0 & 0 \end{bmatrix}$ *and not very interesting.*

The overall system of combined shifts and dilations $\mathcal{L}_2$ *and* $\mathcal{L}_3$ *is*

$$\mathbf{m}(\theta) = \left[\begin{array}{cccc} \theta_0 & \theta_1 & 0 & 0 \\ 0 & \theta_0 & \theta_1 & 0 \\ 0 & 0 & \theta_0 & \theta_1 \\ \hline \theta_0 & 0 & \theta_1 & 0 \\ 0 & \theta_0 & 0 & \theta_1 \\ \hline \theta_0 & 0 & 0 & \theta_1 \end{array}\right]$$

Testing it for identifiability can be done either by a random parameter e.g.

$$\mathbf{m}((1,0)) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

which numerically tests for a one-dimensional annihilator solution (hence $\mathbf{m}$ *is identifiable). Since* $\mathbf{m}$ *is a mosaic family, we can also test its* $\mathcal{M}$*-preserving structures. We have* $\mathbf{m}(\theta) = \theta_0 S_0 + \theta_1 S_1$ *with*

$$S_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Computing the $\mathcal{M}$ *preserving pre-compositions is fairly easy. Let* a_i *denote the rows of* $\mathbf{A}$*.*

- *Compare* $S_0\mathbf{A} = \mathbf{m}(\theta_0, \theta_1)$ *(since* $\mathbf{A} \in \mathcal{S}_{\mathcal{M}}(S_0)$*)*

$$S_0\mathbf{A} = \left[\begin{array}{c} a_1 \\ a_2 \\ a_3 \\ \hline a_1 \\ a_2 \\ \hline a_1 \end{array}\right] = \left[\begin{array}{cccc} \theta_0 & \theta_1 & 0 & 0 \\ 0 & \theta_0 & \theta_1 & 0 \\ 0 & 0 & \theta_0 & \theta_1 \\ \hline \theta_0 & 0 & \theta_1 & 0 \\ 0 & \theta_0 & 0 & \theta_1 \\ \hline \theta_0 & 0 & 0 & \theta_1 \end{array}\right]$$

  yields

$$\mathcal{S}_{\mathcal{M}}(S_0) = \left\{ \begin{bmatrix} \beta & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & \beta & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \right\}$$

- *Similarly,* $S_0\mathbf{A} = \begin{bmatrix} a_2 & a_3 & a_4 & a_2 & a_4 & a_4 \end{bmatrix}^T$ *comparing* $S_1\mathbf{A} = \mathbf{m}(\theta_0, \theta_1)$

(since $\mathbf{A} \in \mathcal{S}_{\mathcal{M}}(S_1)$*) yields*

$$\mathcal{S}_{\mathcal{M}}(S_1) = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & \beta & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \beta \end{bmatrix} \right\}$$

The strong $\mathcal{M}$*-preserving here is the intersection*

$$\mathcal{S}_{\mathcal{M}} = \mathcal{S}_{\mathcal{M}}(S_0) \cap \mathcal{S}_{\mathcal{M}}(S_1)$$

hence $a_{41} = a_{42} = a_{43} = 0$*,* $a_{12} = a_{13} = a_{14} = 0$ *and* $\beta = a_{11} = a_{44}$*, and so*

$$\mathcal{S}_{\mathcal{M}} = \{\lambda \mathtt{Id}\}$$

consists only of scalar matrices, and thereby $\mathcal{M}$ *is potentially identifiable.*

4.6 Summary of Results for FIR Channel Identifiability

We conclude Chapter 4 with the following list of results concerning the ambiguity of sampled multichannel FIR identification:

1. It is sufficient to test the identifiability of a configuration on a single parameter value (following the genericity property).
2. To determine whether a configuration (Σ, Γ, L) is never-identifiable, it is sufficient to test the single-channel case $(\Sigma, \Gamma, 1)$ (either empirically or by combinatorial means), by Theorem 3.15.
3. Configuration parameters affect identifiability as listed in Table 4.2

Table 4.2: Effects of configuration parameters on identifiability

Parameter	**Possible effect on identifiability**	**Remarks**
Domain Σ	Decreases with $\lvert\Sigma\rvert$	If Σ is too small then a solution may be nonexistent
Pattern Γ	Increases with $\lvert\Gamma\rvert$	Decimation results with ambiguity (Theorem 4.10)
Channels L	Increases with L	No improvement beyond $L \geq \lvert\Sigma\rvert$ channels (Theorem 3.15)

CHAPTER 5

ON THE SPATIAL STRUCTURE OF DATA EIGEN-PATCHES

In Chapter 4, output data sample patches of the form

$$X_\tau = \mathtt{D}_\Gamma \varsigma^\tau (\mathbf{h} * x) \tag{5.1}$$

were used to identify the FIR $\mathbf{h}$. That was achieved through computing the space (or rather its ortho-complement) spanned by those patches subject to the annihilator equation (2.13). An orthogonal basis to $\text{span}\{X_\tau\}_{\tau\in\mathcal{E}}$ was computed by solving the eigenvalue problem on the outer products $X_\tau \otimes X_\tau^*$ (equivalently, the SVD of the data matrix), which happens to be known as the *principal components*[1] analysis of the data. It turns out that the principal components of such sampled data have very compelling spatial characteristics.

Assuming that $\mathbf{h} = \delta$, note that (5.1) can be written rather as a restriction

$$X_\tau := \mathtt{D}_\Gamma x_\tau, \qquad \text{where } x_\tau := \varsigma^\tau x \tag{5.2}$$

Here x_τ is merely a shifted version of x.

For the context of this work, principal components of a (finite) patch collection $\{X_\tau\}_{\tau\in\mathcal{E}}$ will be the eigenvectors of the (positive semidefinite) *sample auto-correlation* operator

$$R_\mathcal{E} := \frac{1}{|\mathcal{E}|}\sum_{\tau\in\mathcal{E}} X_\tau \otimes X_\tau^* \tag{5.3}$$

where the scaling by the cardinality $|\mathcal{E}|$ is done for technical reasons. The eigenvectors constitute an orthogonal basis for $\text{span}\{X_\tau\}_{\tau\in\mathcal{E}}$ due to the positive semidefinite nature of $R_\mathcal{E}$ (being a sum of outer products), and can be sorted by their "contribution" weight (see discussion about Rayleigh quotient in [20]).

[1] There are several variations of principal component analysis, most of which fit the vectors into an affine space, and center the data around their mean as a preliminary step. Since we fit the data patches in a *linear* space, the centering step was not performed.

In practice, the analysis is done by "flattening" X_τ averaging over its outer products:

```
# X is of size [sample]x[N1]x[N2]...
num_samples = X.shape[0]
X_flat = reshape(X,(num_samples,-1)) # Flatten other axes
PCA, P_values = eig(X_flat@X_flat.T)
```

After solving the eigenvalue problem the vector are then reshaped into their original tensor form

```
# P_components is of size [data dimension]x[data dimension]
PCA_patches = reshape(PCA,(data dimension, N1, N2...))
```

For example, RGB data patches on a domain would have RGB principal components on the same domain.

In the following examples (see Figures 5.2, 5.3, 5.4, and 5.5), $R_{\mathcal{E}}$ was computed from empirical data from several photos and paintings, on circular and square patches Γ. The offsets $\mathcal{E}$ were scattered in various ways: random, and uniform with overlapping (compare the several modes in Figure 5.1).

The results on those data patches are stellar:

- The principal components of larger eigenvalues resemble harmonic functions on the domain Γ, that is, solutions of the equation

$$\nabla^2 \phi(t) = 0 \qquad\qquad t \in \Gamma \tag{5.4}$$

  Here $\Gamma \subset \mathbb{R}^2$ is a continuous version of the patch domain,

- Furthermore, the spectra of all sampling scenarios are nearly identical.

The aim of this chapter is to give this phenomenon a heuristic explanation.

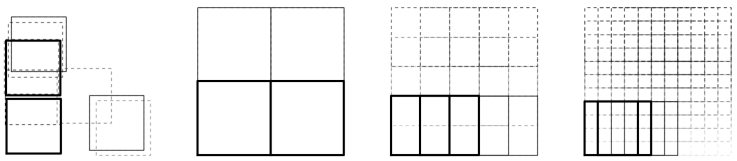


Figure 5.1: Offset scatters (from left to right): random, uniform no overlap, uniform 50% overlap, uniform 75% overlap

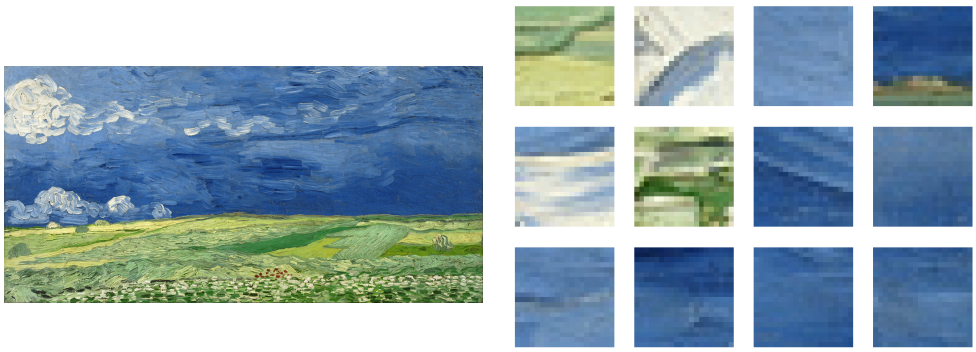


Figure 5.2: Left: *Wheatfield under Thunderclouds* (Van Gogh, 1890), right: a selection of sampled 20×20 patches

Table 5.1: Various PCA results for the Van-Gogh's painting

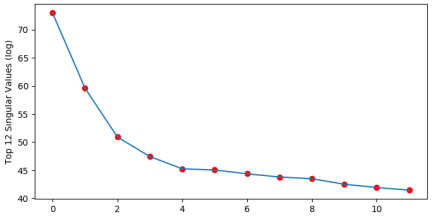
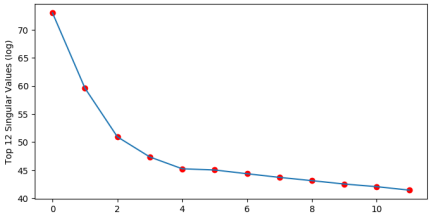
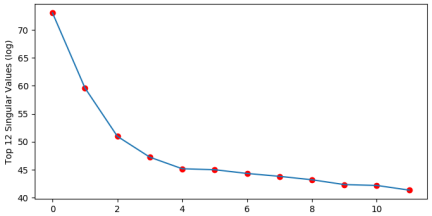
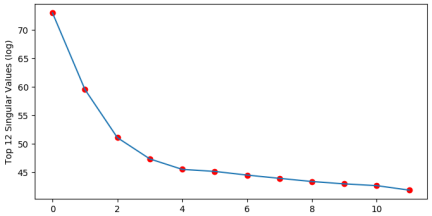
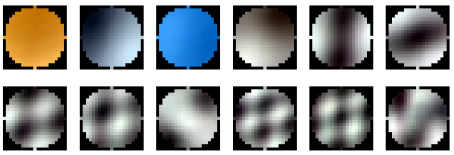
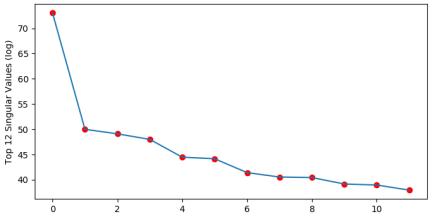
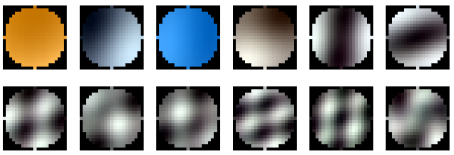
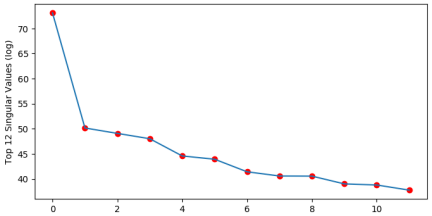
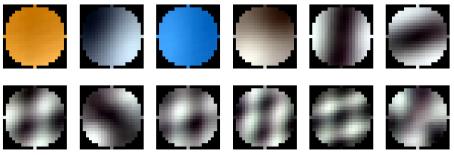
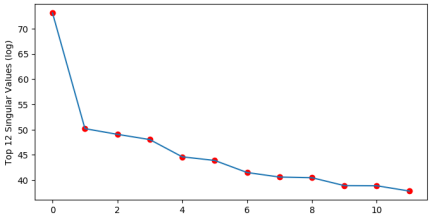
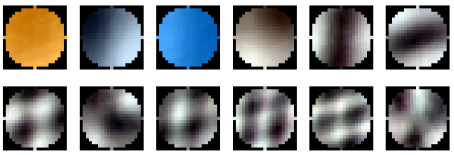
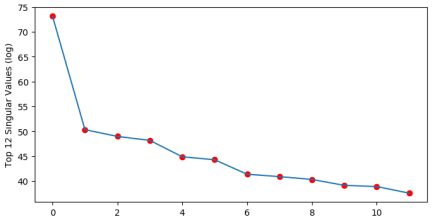
Sampling Pattern	First 12 Principal Components	Singular Values
Random		
Uniform, 75% overlap		
Uniform, 50% overlap		
Uniform, no overlap		

Figure 5.3: Left: *Puerta Del Sol, Madrid* (Elad Yarkony, 2014), right: a selection of sampled circular 20×20 patches

Table 5.2: Various PCA results for Puerta Del Sol photo

Sampling Pattern	First 12 Principal Components	Singular Values
Random		
Uniform, 75% overlap		
Uniform, 50% overlap		
Uniform, no overlap		

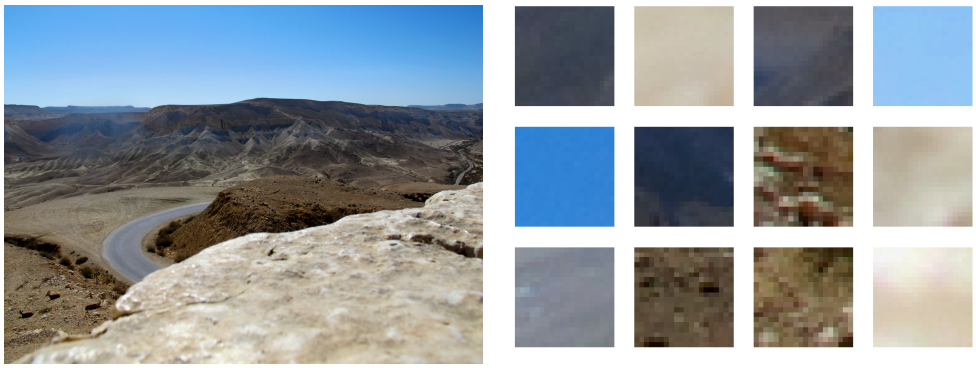
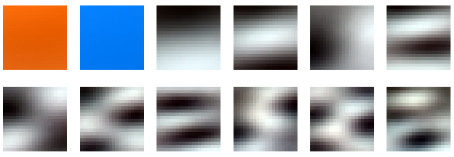
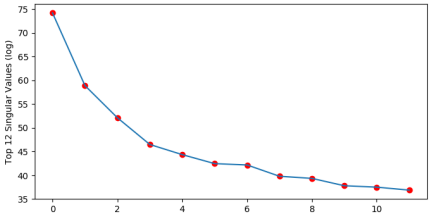
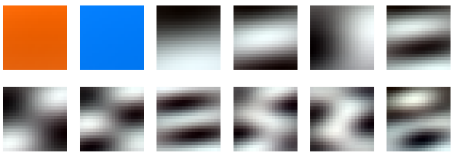
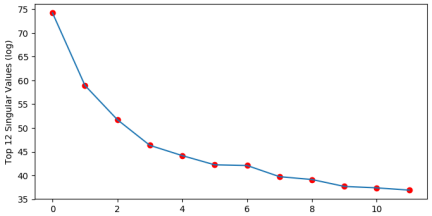
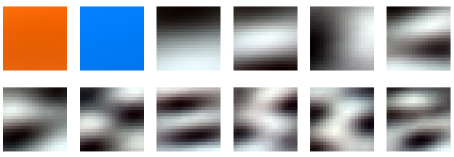
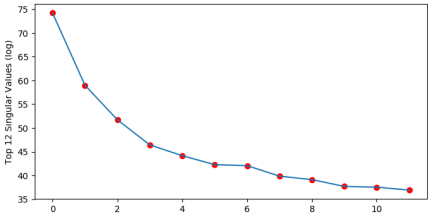
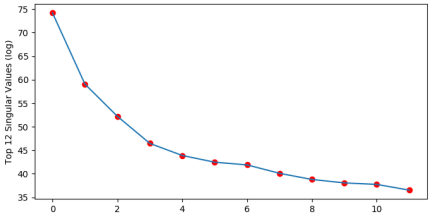


Figure 5.4: Left: *Ramon crater* (Elad Yarkony, 2011), right: a selection of sampled 20×20 patches

Table 5.3: Various PCA results for Ramon crater photo

Sampling Pattern	First 12 Principal Components	Singular Values
Random		
Uniform, 75% overlap		
Uniform, 50% overlap		
Uniform, no overlap		

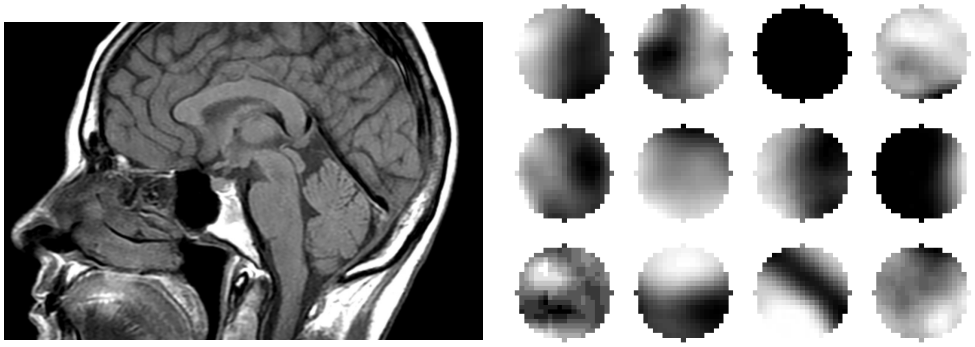
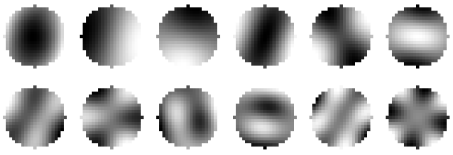
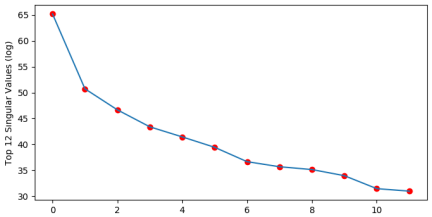
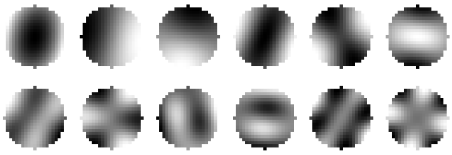
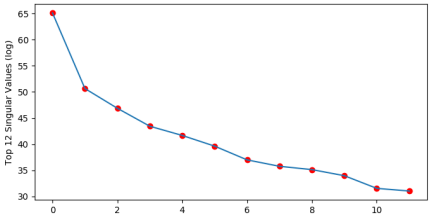
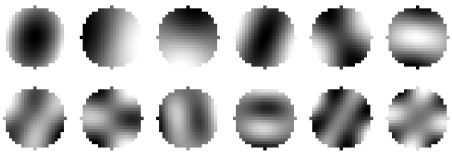
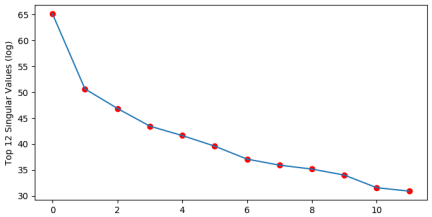
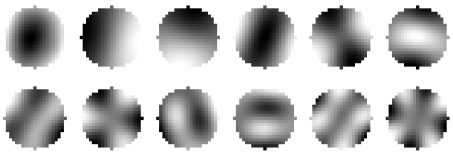
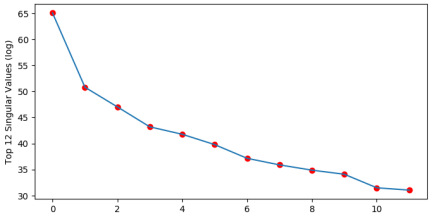


Figure 5.5: Left: MRI image (adapted from [21]), right: a selection of sampled circular patches of radius 10

Table 5.4: Various PCA results for MRI image

Sampling Pattern	First 12 Principal Components	Singular Values
Random		
Uniform, 75% overlap		
Uniform, 50% overlap		
Uniform, no overlap		

This wave-like spatial shape of the principal components alludes to a Toeplitz structure in $R_{\mathcal{E}}$, i.e.

$$\langle s|R_{\mathcal{E}}|t\rangle = \langle s+\tau|R_{\mathcal{E}}|t+\tau\rangle \tag{5.5}$$

for valid index shifts. Even if (5.5) holds by approximation rather than equality, the invariant subspaces $R_{\mathcal{E}}$ would be approximately the invariant subspaces of the nearest Toeplitz operator, by perturbation theory for symmetric matrices (see [22]). Indeed, examining the projection ratio to the space of Toeplitz operators $\mathtt{P}_{tl}(R_{\mathcal{E}})$ (see Table 5.5) is almost exclusively around 1.0.

Table 5.5: Toeplitz score for several experimental results

Source Signal and Sampling Type	Topelitz Score $\frac{\|\mathtt{P}_{tl}(R_{\mathcal{E}})\|}{\|R_{\mathcal{E}}\|}$
Puerta Del Sol, random	0.9999730612861953
Ramon, uniform no overlap	0.9999912466644767
Van Gogh, uniform 1/2 overlap	0.9999868670715982
MRI, uniform 75% overlap	0.9999991347774677

While the Toeplitz structure has been observed in all examples, the autocorrelation kernel $r[t-s]$ is naturally affected by the image itself. An image with prevalent horizontal features would have predominantly horizontal principal components (as apparently, Van Gogh's brush strokes are, see Figure 5.6 and compare with the patches).

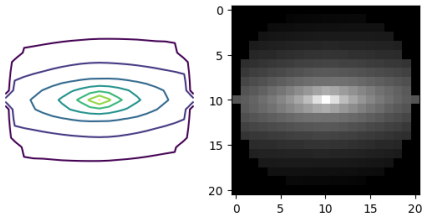


Figure 5.6: Autocorrelation kernel of Van-Gogh's painting

We pose two separate questions:

- Why is the sampled $R_{\mathcal{E}}$ Toepleitz?
- Why do the lower energy eigenvectors of $R_{\mathcal{E}}$ resemble harmonic functions on Γ?

We will try to provide answers the two question in Sections 5.1 and 5.1.2.

5.1 The Toeplitz Nature of $R_{\mathcal{E}}$

To answer why $R_{\mathcal{E}}$ has a Toeplitz kernel, we have to first understand the process from which $R_{\mathcal{E}}$ is computed. The short answer is that for random patch scatter, $R_{\mathcal{E}}$ is approximately (or exactly) Toeplitz, regardless of the signal x, due to the strict stationarity of the signal x_τ where τ is uniformly distributed. For offsets $\mathcal{E}$ that are scattered on a lattice, though, the $R_{\mathcal{E}}$ still exhibits Toeplitz structure, which can be attributed to the signal x itself. We will examine a generative patch model that is wide-sense stationary.

Examining empirical sum $R_{\mathcal{E}}$, shifted on the diagonals yields

$$\langle s+\gamma|R_{\mathcal{E}}|t+\gamma\rangle = \langle s|R_{\mathcal{E}}+\gamma|t\rangle \tag{5.6}$$

An approximated Toeplitz structure means that $R_{\mathcal{E}} \approx R_{\mathcal{E}+\gamma}$ for all γ.

In some cases, this approximation holds regardless of x. For example when $\mathbb{T}$ is a finite group and $\mathcal{E} = \mathbb{T}$, then $\mathcal{E} = \mathcal{E}+\gamma$ and $R_{\mathcal{E}} = R_{\mathcal{E}+\gamma}$. Even if $\mathbb{T}$ is not a finite group, scattering τ uniformly on a large enough box $\mathcal{E}$ will result with an approximated Toeplitz $R_{\mathcal{E}}$.

Assume the index group $\mathbb{T}$ has a shift-invariant measure (a Haar measure), for which $\eta(A+t) = \eta(A)$ for every measurable $A \subset \mathbb{T}$. On a discrete group, it will be the counting measure. This gives a decomposition of $R_{\mathcal{E}}$ to a Toeplitz part plus a residual (see proof in Appendix A).

Lemma 5.1. *We can write $R_{\mathcal{E}}$ as the sum $\langle s|R_{\mathcal{E}}|t\rangle = R(t-s)+\epsilon(t)$, where $\epsilon(t)$ is bounded by*

$$|\epsilon(t)| \le M\frac{\eta((\mathcal{E}+t)\Delta\mathcal{E})}{\eta(\mathcal{E})}$$

where $A\Delta B = A\backslash B + B\backslash A$ is symmetric set-difference, and M is a constant independent on t.

The residue $\epsilon(t)$ quantifies the "boundary" effect of the integration, and in case it is small enough, then $R_{\mathcal{E}}$ is approximately Toeplitz. This is not an uncommon scenario, for example, when $\mathbb{T} = \mathbb{Z}^N$ (or a large discrete torus), $\mathcal{E} = [a,b]^N$ is a cube, and $\Gamma \subset \mathcal{E}$ is a much smaller cube.

On the other extreme, if $\mathcal{E}$ is chosen such that $\{\tau+\Gamma\}$ are pairwise disjoint for different values of τ, then $R_{\mathcal{E}}$ could be any Hermitian matrix. For example, if $\mathbb{T} = \mathbb{Z}^2$ with $\mathcal{E} = \{(10m, 10n), \quad 0 \le m,n < N\}$ and $\Gamma = [0,9]^2$.

For any signal x that is 10 periodic, the outer product $X_\tau \otimes X_\tau^*$ is constant (and not Toeplitz).

More generally, if $\mathcal{E} = \tau_0 + G$ is a coset of some subgroup $G \subset \mathbb{T}$, then $R_\mathcal{E}$ is G-periodic on the diagonals (rather than constant) since $\mathcal{E} + g = \mathcal{E}$ (which is, to some degree, cyclo-stationarity). The examples above, however, show that $R_\mathcal{E}$ is close to Toeplitz *even when the windows do not overlap*, which suggests that the underlying signal x itself is responsible for the structure of $R_\mathcal{E}$. For that we will treat x_τ as a random process (refer to Appendix B.2 for the probabilistic extension of the data model).

The analysis is, nonetheless, easier on compact tori rather than on free index groups, since for the latter we can define a uniform probability distribution on the entire index group. Note that working with finite/compact tori is not just a theoretical relaxation, but also has practical manifestation. Some natural settings are better modeled on tori rather than Euclidean spaces (e.g. optical lenses with spherical coordinates). Also, most problems involving physical signals call for a bounded support, which can be embedded into a compact torus.

On a compact torus $\mathbb{T}$ we define uniform variables as follows:

Definition 5.2. *A mapping $\tau : \Omega \to \mathbb{T}$ is called a* uniform random variable *(denoted $\tau \sim U(\mathbb{T})$) if its probability measure is shift invariant, i.e. $\mathbf{p}(\tau \in A) = \mathbf{p}(t + \tau \in A)$ for all $t \in \mathbb{T}$.*

Since the Haar measure on a compact group $\mathbb{T}$ is unique up to scaling (see [23]), there is a unique uniform probability measure on $\mathbb{T}$, that we will denote by $\mathbf{p}(\tau)$.

The following result is central to this discussion (with surprisingly easy proof, see Appendix A):

Lemma 5.3. *Assuming that $\tau \sim U(\mathbb{T})$ and that x is a random process independent on τ, the process $x_\tau[t] := x[t + \tau]$ is stationary in the strict sense.*

The extent of Lemma 5.3 is not limited for x with numerical values (i.e. complex linear spaces), but rather to x values on in any measurable set. For example, if $\tau \sim U(\mathbb{T})$ is independent on a random mapping $\psi : \mathbb{T} \to \mathbb{T}$, then $\psi(t + \tau)$ is strictly stationary.

Corollary 5.4. *Let x be a random process indexed on $\mathbb{T}$, taking values in a Hilbert space V, and let $\tau \sim U(\mathbb{T})$ be independent on x. Then the process $x_\tau = \varsigma^\tau x$ is stationary, and in particular WSS, with (constant) expectation and autocorrelation*

$$\mathbb{E}(x[t+\tau]) = \int_{\mathbb{T}} \mathbb{E}(x[\tau]|\tau) d\mathbf{p}(\tau) \tag{5.7}$$

$$r[t] := \mathbb{E}(x[t+\tau]x[\tau]) = \int_{\mathbb{T}} \mathbb{E}(x[t+\tau]x[\tau]|\tau) d\mathbf{p}(\tau) \tag{5.8}$$

both are determined by integrating across the torus.

5.1.1 The Statistical Interpretation of $R_\mathcal{E}$

Once we assign a probability model to x_τ, we can examine $R_\mathcal{E}$ from a statistical point of view. The Toeplitz structure indicates that the sampled process $x_\tau[t]$ is wide-sense stationary with respect to the group index t (or at least for shifts in $\Gamma - \Gamma$).

We consider both the case of x deterministic and $\tau \in U(\mathbb{T})$, as well as x WSS and $\tau \in \mathcal{E}$ for a large $\mathcal{E}$. In both cases, we claim that $R_\mathcal{E}$ is an unbiased estimator of $\mathbb{E}(X_\tau \otimes X_\tau^*)$.

- If $x[t]$ is random and WSS and $\mathcal{E}$ is finite then $\mathbb{E}(R_\mathcal{E})$ is Toeplitz:

$$\mathbb{E}(\langle s+\gamma|R_\mathcal{E}|t+\gamma\rangle) = \frac{1}{|\mathcal{E}|}\sum_{\tau\in\mathcal{E}} r[s+\gamma-t-\gamma] = r[s-t]$$

- If $x[t]$ is random or deterministic, and τ is an independent uniform shift on $\mathbb{T}$ (on a large enough subset, that is), then x_τ is strictly stationary, and so $\mathbb{E}(\langle s|R_\mathcal{E}|t\rangle)$ is Toeplitz.

Either way, $R_\mathcal{E}$ is expected to be Toeplitz (or approximately so, if the group is not a compact torus).

We need to weigh in the question of convergence mode and rate, which is left out of this discussion because it veers into ergodic theory. We will mention, however, that for independent samples, than standard strong laws will suffice (and the model x_τ for uniform τ delivers independent samples).

5.1.2 A Generative Image and Patch Models

In the non-random scatter case $R_{\mathcal{E}}$, we will analyze $R_{\mathcal{E}}$ with respect to a generative patch model. Statistical modeling of natural images has been a research topic for a long time (see [24] for a comprehensive survey of models).

We offer a generative model that mimics the appearance of prevalent image patches successively overlaying layers on top of each other (see *dead leaves model* [24]):

$$x_{n+1} = \mathcal{C}(x_n, \xi_n)$$

Here $\mathcal{C}(\cdot, \cdot)$ will be defined below, but it is merely a convex combination of the two layers. This model is paramount in optical image modeling: it is how painters paint, and how 3D graphics engines produce projected images. It can be used to generate individual patches X_τ, or an entire image from which patches are sampled. In both cases, the resulting model is stationary.

To define overlay model, let $f : \mathbb{T} \to V$ be some foreground signal, $m : \mathbb{T} \to [0, 1]$ be an opacity mask (with $m = 1$ fully opaque, $m = 0$ transparent), the complement transparency mask $\bar{m}(t) = 1 - m(t)$, and $\psi : \mathbb{T} \to \mathbb{T}$ be some coordinate transform. Define the composition of f and b by

$$\mathcal{C}(b, f) = b(t) \cdot \bar{m}(\psi(t)) + f(\psi(t)) \cdot m(\psi(t)) \tag{5.9}$$

in which $f(\psi(t))$ overrides b on $m(\psi(t))$ (see Figure 5.7).

Figure 5.7: Dead leaves model (layer occlusion)

The interesting thing about this model is its ability to produce stationary signals, as long as the layers are shifted randomly i.e. $\psi(t) = \phi(t + \tau)$ for some uniform shift $\tau \in U(\mathbb{T})$:

Lemma 5.5. *If the following hold:*

1. *b and f are independent processes indexed on $\mathbb{T}$ (representing background and foreground signals respectively).*
2. *b is wide-sense stationary.*
3. *$\psi(t) = \phi(t + \tau)$ where $\tau \sim U(\mathbb{T})$ and $\phi : \mathbb{T} \to \mathbb{T}$ is random and independent on τ.*

then $\mathcal{C}(b, f)$ as defined in (5.9) is wide-sense stationary.

To conclude, one can generate a WSS signal by summing or overlaying WSS components with random shifts on $\mathbb{T}$. The remaining question is what kind of (WSS) elements ξ_n one can throw into this construction process, which can include virtually any planar shape filled with deterministic or random textures.

One reasonable option is to take some shape (with some filling) rotate and shift it on the domain, or in other words, perform a random rigid transformation (see Figure 5.8 for example). Affine-transformed template is anything of the form

$$\xi_n(t) = f(R(t + \tau)) \tag{5.10}$$

where R is a random matrix, $\tau \sim U(\mathbb{T})$, and f is either deterministic or random, where all random quantities are independent. The random shift guarantees stationarity (due to Lemma 5.3).

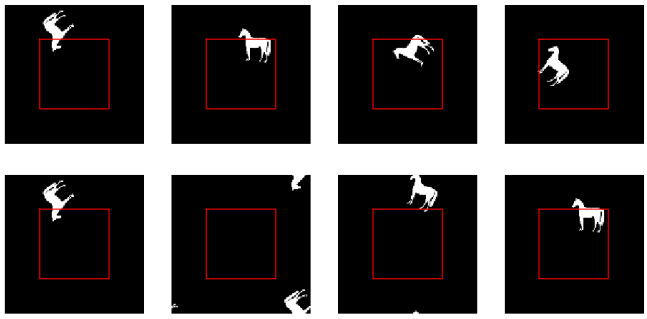


Figure 5.8: Affine-transformed template

In practice, rotations are performed in $\mathbb{R}^D$ and mapped to the torus by a quotient map. Unfortunately, rotations are not well behaved through the quotient map (they are not invertible or associative as they would be on $\mathbb{R}^d$). If, however, we restrict our discussion to the unit ball $\mathcal{B}([-1,1]^D)$ which is bounded by the flat torus $[-1,1]^D$, then we can safely write $R_1(R_2x) = (R_1R_2)x$ and $RR^*x = R^*Rx = x$ for all $x \in \mathcal{B}([-1,1]^D)$.

An interesting observation about the model (5.10) is its radially symmetry (it is easy to verify that every composition of radially symmetric layers is radially symmetric as well). Empirically, it is apparent that the autocorrelation kernel of the rotated horse template (Figure 5.8) depends on the *distance* between indices $\mathbb{E}(x[t]x[s]) \approx r(d(t,s))$ rather than absolute locations s,t or relative location $s-t \in \mathbb{T}$. The autocorrelation as a function of the distance is shown in Figure 5.9.

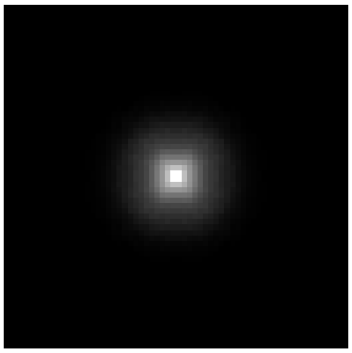


Figure 5.9: Autocorrelation of rigid-transformed template in Figure 5.8

We remark here that $d(s,t)$ is well defined on the torus $[-1,1]^D$ by geodesics, but for the sake of this discussion we assume that $s,t \in \Gamma$ where Γ is small enough to approximate $d(s,t) \approx \|s-t\|$ here $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^D$.

Modeling such processes with radially symmetric autocorrelation calls for the notion of *locally rotation invariant* processes on $\mathbb{T}$.

Definition 5.6. *We say that a process $x : \mathbb{T} \to V$ is rotation invariant (locally, on Γ) if the joint probability distribution of x is fixed under rotations, i.e. for every finite collection $K \subset \Gamma$ one has*

$$\mathbf{p}(\mathtt{D}_K x \in A) = \mathbf{p}(\mathtt{D}_{QK} x \in A)$$

Here Q is a rotation on $[-1,1]^D$, and $\mathtt{D}$ is the sampling/restriction operator.

Informally, it means that the joint statistics of x, when sampled on a small enough patch (able to rotate freely in the torus) is invariant under rotations.

Lemma 5.7. *Let y be a random process defined on $[-1,1]^D$ supported on $\Gamma \subset \mathcal{B}(\mathbb{R}^D)$ (note that rotating Γ keeps it inside $[-1,1]^D$). Then the process $x(t) = y(Rt)$ where $R \in \mathbb{R}^{D\times D}$ is a rotation drawn from uniform distribution is locally radially symmetric on Γ.*

Therefore, composing a uniform rotation to any process makes it (locally) isotropic, and so $y(Rt)$ (the un-shifted template) has distributions that are invariant to rotations.

Lemma 5.8. *Let $y(t)$ be locally radially symmetric on Γ. Then the (stationary) process $x(t) = y(t+\tau)$ has an autocorrelation kernel that is locally isotropic.*

Proof. Due to the stationarity, we have $\mathbb{E}(x(t)x(s)) = \mathbb{E}(x(t-s)x(0)) = r(s-t)$. Nevertheless, the joint distribution of $x(t-s)$ and $x(0)$ is constant under rotations, and so the expectation $r(t-s)$ is also constant under rotations:

$$\begin{aligned} r(t-s) &= \int x(t-s)x(0)d\mathbf{p}(\omega) \\ &= \int_\Omega x(Q(t-s))x(Q0)d\mathbf{p}(\omega) \\ &= r(Q(t-s)) \end{aligned}$$

The latter is obtained by invoking Lemma 5.7 with $K = \{0, t-s\}$, which completes the proof. □

To complete the discussion, we note that whenever r is restricted to Γ where $d(s,t) \approx \|s-t\|$, there exists some function $\rho : \mathbb{R}^+ \to \mathbb{C}$ such that

$$r(s-t) \approx \rho(\|s-t\|)$$

where $\rho(d) = r(d\hat{e})$ for any unit vector $\hat{e} \in \mathcal{B}(\mathbb{R}^D) \subset [-1,1]^D$.

There is a bound on the autocorrelation for generative models. Let $\beta_{\mathcal{R}}(d)$ denote the cap volume of the canonical $\mathcal{R}$-radius ball in $\mathbb{R}^d$ intersected with the plane $x_1 = d$, which is generally a decreasing function with $\beta_{\mathcal{R}}(0)$ maximal value (and equals half of the overall volume), and $\beta_{\mathcal{R}}(2\mathcal{R}) = 0$.

Lemma 5.9. *Assume that y is deterministic and bounded by M, whose support is bounded by $B = \mathcal{B}_{\mathcal{R}}([-1,1]^D)$ with $\mathcal{R} \leq \frac{1}{2}$. Then the process*

$x(t) = y(R(t+\tau))$ is WSS with radial autocorrelation bounded by

$$|r(t)| \leq 2M^2 \beta_{\mathcal{R}}(\|t\|) \tag{5.11}$$

For example, on the two-dimensional torus $[-1,1]^2$ we have

$$|r(t)| \leq 2M^2 \left[\mathcal{R}^2 \arccos(\frac{\|r\|}{2\mathcal{R}}) - \frac{1}{2}\|t\|\sqrt{4\mathcal{R}^2 - \|t\|^2} \right]$$

This is a useful measure of how two neighboring pixels are correlated depending on their distance.

5.2 The Eigenvalue Problem of Symmetric Toeplitz Operators

The first few eigen-patches of $R_{\mathcal{E}}$ look a lot like low-energy solutions of a harmonic problem on the patch domain. Those are solutions of an eigenvalue problem of a Toeplitz and symmetric R:

$$R\phi = \lambda\phi \tag{5.12}$$

where $\phi \in \mathbb{C}^{\Gamma} \otimes \mathbb{C}^{L}$. If we can embed Γ in a sufficiently large group $\mathbb{T}$, R can be written in terms of convolution on $\mathbb{T}$ followed by restriction: $R = \mathtt{D}_{\Gamma}\mathtt{C}_{r}\mathtt{D}_{\Gamma}^{*}$, where $r[\tau] = \langle \tau | R_{\mathcal{E}} | 0 \rangle$ for $\tau \in \Gamma - \Gamma$ (differences taken on $\mathbb{T}$), and padded with zeros elsewhere. Thus, the eigenvalue problem in R can be written as a generalized eigenvalue problem augmented on $\mathbb{T}$:

$$r * \psi = \lambda \mathtt{P}_{\Gamma}\psi \tag{5.13}$$

Here $\psi \in \mathbb{C}^{\mathbb{T}} \otimes \mathbb{C}^{L}$.

Unfortunately, no general solution for (5.13) has been offered (to this date). In the one-dimensional case, it has been shown that discrete cosine transform can approximately diagonalize R, see e.g. [25], which belongs in a larger class of solutions relying on circulant approximations.

One thing to note is that R has the structure of a difference equation with boundary condition. Let Θ denote the (symmetric) support of $r[s]$. By the symmetry $r[s] = r[-s]$ we can write the autocorrelation as

$$R = \frac{1}{2}\sum_{s\in\Theta} r[s]\mathtt{D}_{\Gamma}(\varsigma^{s} + \varsigma^{-s})\mathtt{D}_{\Gamma}^{*} = \frac{1}{2}\sum_{s\in\Theta} r[s]L_{s}$$

where $L_s = \mathtt{D}_\Gamma(\sigma^s + \sigma^{-s})\mathtt{D}_\Gamma^*$ is the symmetric truncated Laplace operator of order s. The equation $R\phi = \lambda\phi$ can be broken into a difference equation on the erosion set

$$\sum_{s\in\Theta} r[s](\phi[t+s] + \phi[t-s]) = \lambda\psi[t] \qquad\qquad t \in \Gamma \ominus \Theta$$

with the boundary condition zeroing $\phi[t \pm s]$ whenever $t \notin \Gamma \ominus \Theta$.

The analogy to Laplace partial difference equation ends at this point, since L_s cannot be written as a polynomial of shifts, and approximating L_s that way has failed. To illustrate the departure from difference equations, consider the one-dimensional case $\mathbb{C}^N$, $L_s = J + J^*$, where J is a Jordan block with 1 on the super-diagonal, we have $L_s^2 = J^2 + (J^*)^2 + 2(JJ^*)$. The matrix JJ^* is diagonal with 1 on all elements except the first and last, setting it apart from the identity, therefore $L_s^2 \neq J^2 + (J^*)^2 + \mathtt{Id}$, and so eigenvectors of L_s and L_s^2 are not necessarily the same (and the difference will grow with higher powers of L_s).

Another potential direction (that has not been fully exploited) is to compare the quadratic form $R_\mathcal{E}$ with a quadratic form of a harmonic problem, which has very similar waveform eigenvectors. Assume that Γ is a discrete lattice on $\mathbb{R}^2$, and $\tilde{\Gamma} \subset \mathbb{R}^2$ is a simply connected domain containing Γ. Define the quadratic form associated with the positive definite kernel r by

$$\mathbb{Q}(\phi) = \langle R\phi, \phi\rangle = \sum_{s,t\in\Gamma} r[s-t]\phi(s)\phi(t)$$

(this samples ϕ on the lattice). Also, define the quadratic form

$$\mathcal{Q}(\phi) = \int_{\tilde{\Gamma}} \|\nabla\phi\|^2 dt$$

The unit eigenvectors $\phi_1, \ldots, \phi_k$ of $\mathcal{Q}$, corresponding to the lowest eigenvalues $\mathcal{Q}(\phi_k)$ (i.e. low-frequency sinusoids), were empirically found to be the highest eigenvectors of the $\mathbb{Q}$. This means, quantitatively, that $\mathbb{Q}$ and $\mathcal{Q}$ share common eigenvectors ϕ_k with reciprocal eigenvalues, i.e.

$$\mathbb{Q}(\phi_k)\mathcal{Q}(\phi_k) \approx c_k$$

This similarity is likely due to the properties of the kernel r, at least when it is a radially-symmetric decreasing function. This direction requires further study, and makes a possible future development of this research.

APPENDIX A

PROOFS

A.1 Proofs for Chapter 2

Lemma 2.13

Proof. Since X is irreducible, so is the Cartesian square X^2. Define the set of all ambiguous parameter pairs:

$$R = \left\{ (\theta, \theta') \;\middle|\; p(\theta) = p(\theta') \right\} \subset X^2 \tag{A.1}$$

containing all pairs drawn from X that are mapped to the same value of p (hence indistinguishable). The diagonal of all the pairs (θ, θ),

$$\Delta = \{(\theta, \theta) \;\mid\; \theta \in X\} \subset R \tag{A.2}$$

naturally embedded into R (as clearly $p(\theta) = p(\theta)$), and should be removed from it. We are left with the complement set

$$R_p := R \setminus \Delta \subseteq R \subset X^2 \tag{A.3}$$

The set $X_=$ of all parameters with nontrivial ambiguous pairings can be written as

$$X_= := \pi(R_p) \tag{A.4}$$

where $\pi : X^2 \to X$ is the projection to the first factor

$$\pi(\theta_1, \theta_2) := \theta_1 \tag{A.5}$$

The Zarisky closure $\overline{X_=} \subset X$ is algebraic (see [18]), which is either equal to X (in which case $X_=$ is dense in X), or nowhere dense in X (in which case $X_=$ is nowhere dense in X), which concludes the proof for $X_=$.

For the proportional ambiguity claim on $X_\propto$ define the polynomial mapping $q(\theta_1, \theta_2) : X^2 \to \mathbb{C}^{n(n-1)/2}$ taking 2×2 determinantal values

$$\begin{vmatrix} p_k(\theta_1) & p_k(\theta_2) \\ p_j(\theta_1) & p_j(\theta_2) \end{vmatrix}$$

which is also the exterior product $p(\theta_1) \wedge p(\theta_2)$. Then, having $X_\propto = \pi(R_q)$ the rest of the proof is the same.

□

Lemma 2.20

Proof.

- For parts (1) and (2), if the range of $\tilde{\mathbf{m}}$ is the same for θ and $\tilde{\theta}$, i.e. $\mathtt{Im}(T\mathbf{m}(\theta)S) = \mathtt{Im}(T\mathbf{m}(\tilde{\theta})S)$, then

  $$T\mathbf{m}(\theta)S = T\mathbf{m}(\tilde{\theta})SK$$

  for some $K \in \mathtt{Lin}(\tilde{U})$. Applying the right and left inverses of S and T respectively, results in

  $$\mathbf{m}(\theta) = \mathbf{m}(\tilde{\theta})SKS^\dagger$$

  so that $\boldsymbol{\rho}(\theta) \subset \boldsymbol{\rho}(\tilde{\theta})$, and by symmetry we can show that $\boldsymbol{\rho}(\theta) \supset \boldsymbol{\rho}(\tilde{\theta})$, and deduce (as $\mathbf{m}$ is range-identifiable) that $\theta = \tilde{\theta}$. Note that $\mathbf{m}$ and $\tilde{\mathbf{m}}$ maintain the same rank stratification in $\mathbf{M}$, as $\mathrm{rk}(\mathbf{m}) = \mathrm{rk}(\tilde{\mathbf{m}})$.

- Part (3) is immediate, as $\rho = \rho \circ f \circ f^{-1}$.

□

Lemma 2.27

Proof. Consider the subspace given by intersection

$$W_s = \bigcap_{\theta \in \mathbf{M}} \mathtt{Ker}(\mathbf{m}(\theta)) = \bigcap_{i=1}^{n} \mathtt{Ker}(S_i) \subset U$$

containing all nullvectors in U common to all θ. Let S be comprised of columns that are orthonormal basis vectors for $W_s^\perp$ (such that if $W = \{0\}$,

then S is just identity). Similarly, let

$$W_t = \bigcap_{\theta \in \mathbf{M}} \mathtt{Ker}(\mathbf{m}(\theta)^*) = \bigcap_{i=1}^{n} \mathtt{Ker}(S_i^*) \subset V$$

be the set of all common null co-vectors, and likewise define T whose columns are an orthonormal basis for $W_t^\perp$. As SS^* and TT^* are projections on the common domain and co-domain, we have

$$\mathbf{m}(\theta) = T \underbrace{T^* \mathbf{m}(\theta) S}_{\tilde{\mathbf{m}}(\theta)} S^*$$

The mapping $\tilde{\mathbf{m}}(\theta) := T^* \mathbf{m}(\theta) S \in \mathtt{Lin}(\mathrm{colsp}(S), \mathrm{colsp}(T))$ has full typical rank, for otherwise there would be a linear combination of the basis elements $\{S_i\}$ that has left and right nullvectors left out of S and T. □

A.2 Proofs for Chapter 3

Proposition 3.5

Proof. The leftmost inclusion $\mathcal{M}(\alpha \mathtt{Id}) \subset \mathcal{M}$ is trivial ($S(\alpha \mathtt{Id}) = \alpha S$). Next, if $\mathcal{M}\mathbf{A} \subset \mathcal{M}$, then $\mathcal{M}'\mathbf{A} \subset \mathcal{M}$ for every subset $\mathcal{M}' \subset \mathcal{M}$, proving $\mathcal{S}_\mathcal{M} \subset \mathcal{S}_\mathcal{M}(\mathcal{M}')$. Finally, for $\mathbf{A}$ that satisfies $\mathcal{M}'\mathbf{A} \subset \mathcal{M}$, also satisfies $S\mathbf{A} \in \mathcal{M}$ for all $S \in \mathcal{M}'$, proving that $\mathcal{S}_\mathcal{M}(\mathcal{M}') \subset \mathcal{S}_\mathcal{M}(S)$ for every S. □

Proposition 3.9

Proof. We have $\mathbf{A} \in \bigcap_{i=1}^n \mathcal{S}_{\mathcal{M}_i}(S_i)$, iff

$$S_i \mathbf{A} \in \mathcal{M}_i \qquad \text{for all } 1 \le i \le n$$

or $S\mathbf{A} \in \mathcal{M}_1 \times \ldots \times \mathcal{M}_n$, or $\mathbf{A} \in \mathcal{S}_{\mathcal{M}_1 \times \ldots \times \mathcal{M}_n}(S)$, which concludes the proof. □

Proposition 3.10

Proof. For the $\boxed{\subset}$ direction, let $\mathbf{A} \in \mathcal{S}_\mathcal{M}(\mathcal{M}')$, so clearly any linear combination of representatives of $\mathcal{M}_i$ is sent by $\mathbf{A}$ to $\mathcal{M}$. In particular $\mathcal{M}_i\mathbf{A} \subset \mathcal{M}$ for every $1 \le i \le n$, which proves the $\boxed{\subset}$ inclusion as required.

For the $\boxed{\supset}$ direction, let $\mathbf{A} \in \bigcap_{i=1}^{n} \mathcal{S}_{\mathcal{M}}(\mathcal{M}_i)$ so that $\mathcal{M}_i\mathbf{A} \subset \mathcal{M}$ for all $i = 1, \ldots, n$. Every combination in $\mathcal{M}'$

$$S = \sum_{i=1}^{n} \alpha_i S_i \in \mathcal{M}'$$

(where $S_i \in \mathcal{M}_i$) has $S_i\mathbf{A} \in \mathcal{M}$, and overall

$$S\mathbf{A} = \sum_{i=1}^{n} \alpha_i \underbrace{S_i\mathbf{A}}_{\in\mathcal{M}} \in \mathcal{M}$$

which concludes the proof. □

Proposition 3.11

Proof.

1. Let $\mathcal{M}' \subset \mathcal{M}$ and $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{S}(\mathcal{M}')$, then

   $$\mathcal{M}'(\alpha\mathbf{A}_1 + \beta\mathbf{A}_2) = \mathcal{M} + \mathcal{M} = \mathcal{M}$$

   so that $\mathcal{S}_{\mathcal{M}}(\mathcal{M}')$ is a linear space. As a special case, we have $\mathcal{S}_{\mathcal{M}}(S) = \mathcal{S}_{\mathcal{M}}(\text{span}(S))$.

2. From the last part we established that $\mathcal{S}_{\mathcal{M}}$ is a linear space (by taking $\mathcal{M}' = \mathcal{M}$). For multiplication, every $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{M}$ satisfy

   $$\mathcal{M}\mathbf{A}_1\mathbf{A}_2 \subset \mathcal{M}\mathbf{A}_2 \subset \mathcal{M}$$

   which completes the proof.

□

Proposition 3.14

Proof. Let $\mathbf{A} \in \mathcal{S}_{\mathcal{M}}$. For every $S \in \mathcal{M}^I$ we have

$$(S\mathbf{A})_i = \underbrace{S_i\mathbf{A}}_{\in\mathcal{M}} \text{ for all } i \in I$$

so that $S\mathbf{A} \in \mathcal{M}^I$, and then $\mathcal{M}^I\mathbf{A} \subset \mathcal{M}^I$, or $\mathbf{A} \in \mathcal{S}_{\mathcal{M}^I}$, proving that $\mathcal{S}_{\mathcal{M}^I} \supset \mathcal{S}_{\mathcal{M}}$. Conversely, if $\mathbf{A} \in \mathcal{S}_{\mathcal{M}^I}$, then for every $S \in \mathcal{M}^I$ we have

$$S\mathbf{A} \in \mathcal{M}^I$$

Let $S_0 \in \mathcal{M}$, and define $S_{i_0} = S_0$ and $S_i := 0$ for all other $i \neq i_0$. Since $S\mathbf{A} \in \mathcal{M}^I$, then $S\mathbf{A} \in \mathcal{M}$, hence $\mathcal{M}\mathbf{A} \subset \mathcal{M}$ which proves the opposite inclusion. □

Theorem 3.15

Proof. A generic $S \in \mathcal{M}^I$ has linearly *independent* components S_i, hence $\operatorname{span}\{S_i\}_{i\in I} = \mathcal{M}$ (i.e. each channel corresponds to a basis vector of $\mathcal{M}$). Next, let $\mathbf{A} \in \mathcal{S}_{\mathcal{M}^I}(S)$, that is, $S\mathbf{A} \in \mathcal{M}^I$, or $S_i\mathbf{A} \in \mathcal{M}$ for all $i \in I$. Then

$$\mathbf{A} \in \cap_{i\in I}\mathcal{S}_{\mathcal{M}}(S_i\mathbf{A}) = \mathcal{S}_{\mathcal{M}}(\operatorname{span}\{S_i\}_{i\in I}) = \mathcal{S}_{\mathcal{M}}(\mathcal{M}) = \mathcal{S}_{\mathcal{M}}$$

or $\mathcal{S}_{\mathcal{M}^I}(S) \subset \mathcal{S}_{\mathcal{M}}$. The converse is immediate

$$\mathcal{S}_{\mathcal{M}} = \mathcal{S}_{\mathcal{M}^I} \subset \mathcal{S}_{\mathcal{M}^I}(S)$$

and we prove the lemma by mutual inclusion. □

A.3 Proofs for Chapter 4

Theorem 4.2

Proof. 1. Whenever $\Omega \subset \tilde{\Omega}$ we have $\mathtt{D}_\Omega = \mathtt{D}_\Omega\mathtt{D}_{\tilde{\Omega}}$ so $\mathbf{m}$ has the factorization

$$\mathbf{m}(\theta) = \mathtt{D}_\Gamma C_{\mathbf{h}} = \mathtt{D}_\Gamma\mathtt{D}_{\tilde{\Gamma}}\mathtt{C}_{\mathbf{h}} = \mathtt{D}_\Gamma\tilde{\mathbf{m}}(\theta)$$

Since $\mathtt{D}_\Gamma$ is surjective and $\mathbf{m}(\theta)$ is identifiable, then $\tilde{\mathbf{m}}(\theta)$ is identifiable according to Theorem 2.20. The corresponding family has the same parameter space $\mathbf{M}$, but the domain and co-domain become $\tilde{U} = \left\langle \tilde{\Gamma} - \Sigma \right\rangle$, $\tilde{V} = \mathbf{k}^L \otimes \left\langle \tilde{\Gamma} \right\rangle$.

2. The parameter space $\tilde{\mathbf{M}} = \mathbf{k}^L \otimes \mathbf{k}^{\tau+\Sigma}$ is isomorphic to $\mathbf{M} = \mathbf{k}^L \otimes \mathbf{k}^\Sigma$ by mere shift. The domain and co-domain spaces are $\tilde{U} = \langle \gamma + \Gamma - \Sigma - \tau \rangle$, $\tilde{V} = \mathbf{k}^L \otimes \langle \gamma + \Gamma \rangle$. Let $(\varsigma^\tau x)[t] := x[t - \tau]$ denote the shift operator on $\mathbf{k}^{\mathbb{T}}$ (which extends to any subspaces and tensor products). If $\mathbf{h} \in \mathbf{k}^L \otimes \langle \Sigma \rangle$, then $\varsigma^\tau \mathbf{h} \in \mathbf{k}^L \otimes \langle \tau + \Sigma \rangle$. Using the identity

$$\mathtt{D}_{t+\Omega} = \varsigma^t \mathtt{D}_\Omega \varsigma^{-t} \tag{A.6}$$

(shift, sample, and shift back) one can write

$$\begin{aligned}\tilde{\mathbf{m}}(\theta) &= \mathtt{D}_{\gamma+\Gamma}\mathtt{C}_{\varsigma^\tau\mathbf{h}}\mathtt{D}_{\gamma-\tau+\Gamma-\Sigma} = \varsigma^\gamma \mathtt{D}_\Gamma \sigma^{-\gamma} C_{\varsigma^\tau\mathbf{h}} D_{\gamma-\tau+\Gamma-\Sigma} \\ &= \varsigma^\gamma \mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} \varsigma^{\tau-\gamma} D_{\gamma-\tau+\Gamma-\Sigma} \\ &= \sigma^\gamma \mathtt{D}_\Gamma \mathtt{C}_\mathbf{h} \mathtt{D}_{\Gamma-\Sigma} \varsigma^{\gamma-\tau} = \varsigma^\gamma \mathbf{m}(\theta) \varsigma^{\gamma-\tau}\end{aligned}$$

so that $\tilde{\mathbf{m}}$ is a composition of an identifiable $\mathbf{m}(\theta)$ with invertible linear maps, and must be identifiable due to Theorem 2.20.

3. For the group automorphism, the convolutions writes

$$(\mathbf{h} * x) \circ \Phi = (\mathbf{h} \circ \Phi) * (x \circ \Phi)$$

so

$$\mathtt{D}_{\Phi\Gamma}\mathtt{C}_\mathbf{h} = \mathtt{C}_{T_\Phi \mathbf{h}} T_\Phi$$

Here $T_\phi x := x \circ \phi$. Consider the parametric model:

$$\tilde{\mathbf{m}}(\theta) := \mathtt{D}_{\Phi\Gamma}\mathtt{C}_{\mathbf{h}\circ\Phi^{-1}} = \mathtt{D}_\Gamma \mathtt{C}_{T_\Phi \mathbf{h}} \mathbf{h} T_\Phi$$

which is, again, identifiable due to Theorem 2.20.

□

Proposition 4.5

Proof. Firstly, sampled FIR operator $S_\mathbf{h}$ has the form of a sum of indicators

$$S_\mathbf{h} = \sum_{\sigma\in\Sigma}\sum_{\gamma\in\Gamma} \mathbf{h}[\sigma] \left|\gamma\right\rangle \left\langle\gamma-\sigma\right| = \sum_{\sigma\in\Sigma} \mathbf{h}[\sigma] S_\sigma$$

where $S_\sigma := \sum_{\gamma\in\Gamma} \left|\gamma\right\rangle \left\langle\gamma-\sigma\right|$ are indicators, making it a mosaic on the sets

$$J_\sigma := \{(\gamma, \gamma-\sigma)\} \subset \mathcal{V}\times\mathcal{U}$$

This is also a admissible mosaic:

- For $\sigma \neq \sigma'$ we have $J_{\sigma'} \cap J_\sigma = \varnothing$.
- The mapping $v := \gamma \mapsto u := \gamma - \sigma$ is an injective function.
- The union over all $\bigsqcup_{\sigma\in\Sigma} J_\sigma$ covers all possible coordinates in $\mathcal{U}$.

□

Lemma 4.6

Proof. From (4.14) we define the sets $\{\mathcal{U}_{\sigma,\sigma'}\}$ and $\bar{\mathcal{U}}$ on $\mathcal{U}\times\mathcal{U}$:

$$\mathcal{U}_{\sigma,\sigma'} = \big\{\big(u_{\sigma'}(v), u_{\sigma}(v)\big) \mid v \in \mathcal{V}\big\} \tag{A.7}$$

$$\mathcal{U}_0 = \big\{\big(u_{\sigma}(v), u\big) \mid u \in \mathcal{U},\ v \in \mathcal{V},\ \sigma \in \mathbf{\Sigma}\big\} \tag{A.8}$$

Due to (4.14) and (4.15), the value $A_{u,u'}$ is constant along $(u,u') \in \mathcal{U}_{\sigma,\sigma'}$ or $(u,u') \in \mathcal{U}_0$. However, those sets in (A.7) may intersect with each other or with (A.8). Then the values of A should be constant on any collection of intersecting tiles.

Our next step is to define an undirected graph $\mathcal{G} = (\{0\} \cup \mathbf{\Sigma}^2, E)$. The nodes of this graph are pronounced connected if whenever they intersect:

$$\{(\sigma_1,\sigma'_1),(\sigma_2,\sigma'_2)\} \in E \qquad \text{if} \quad \mathcal{U}_{\sigma_1,\sigma'_1} \cap \mathcal{U}_{\sigma_2,\sigma'_2} \neq \varnothing$$

$$\{(\sigma,\sigma'),0\} \in E \qquad \text{if} \quad \mathcal{U}_{\sigma,\sigma'} \cap \mathcal{U}_0 \neq \varnothing$$

Let $\{C_k\}$ be the connected components of $\mathcal{G}$, containing all connected pairs, and let C_0 be the class connected to $\mathcal{U}_0$. We define the mosaic $\{\tilde{J}_k\}$ on $\mathcal{U}\times\mathcal{U}$ by those connectivity classes of $\mathcal{G}$:

$$\tilde{J}_k := \bigcup_{(\sigma,\sigma')\in C_k} \mathcal{U}_{\sigma,\sigma'}$$

Subject to this definition, the value of $\mathbf{A}_{u,u'}$ is constant for all (u,u') belonging to connected components of $\mathcal{G}$, and vanish on $\tilde{J}_0$, that is,

$$\mathbf{A}_{u,u'} = \mathbf{A}_{u,u''} \qquad (u,u'),(u,u'') \in \tilde{J}_k$$

$$\mathbf{A}_{u,u'} = 0 \qquad (u,u') \in \tilde{J}_0$$

which is a mosaic as required. □

Theorem 4.7

In order to prove Theorem 4.7, we first state the following lemma:

Lemma A.1. *Let $\{J_\sigma\}$ be a FIR mosaic. If two pre-composition tiles $\mathcal{U}_{\sigma_1,\sigma'_1}$ and $\mathcal{U}_{\sigma_2,\sigma'_2}$ (as defined in Equation* (A.7)*) are connected (i.e. intersect), then $\sigma_1 - \sigma_2 = \sigma'_1 - \sigma'_2$.*

Proof. (Lemma A.1) We first assume two neighboring (intersecting) $\mathcal{U}_{\sigma_1,\sigma'_1}$

and $\mathcal{U}_{\sigma_2,\sigma'_2}$. In this case, there exists some $\gamma_1, \gamma_2 \in \mathcal{V}$ such that

$$(\gamma_1 - \sigma_1, \gamma_1 - \sigma'_1) = (\gamma_2 - \sigma_2, \gamma_2 - \sigma'_2)$$

so each coordinate equates:

$$\gamma_1 - \sigma_1 = \gamma_2 - \sigma_2 \qquad\qquad \gamma_1 - \sigma'_1 = \gamma_2 - \sigma'_2$$

from which we deduce (by subtracting indices) $\sigma_1 - \sigma_2 = \sigma'_1 - \sigma'_2$ as required. Every two connected sets have a path $\{(\sigma_i, \sigma'_i), \quad i = 1, \ldots, N\}$, such that every two consecutive nodes are intersecting, hence:

$$\begin{aligned}
\sigma_1 - \sigma_2 &= \sigma'_1 - \sigma'_2 \\
&\vdots \\
\sigma_i - \sigma_{i+1} &= \sigma'_i - \sigma'_{i+1} \\
&\vdots \\
\sigma_{N-1} - \sigma_N &= \sigma'_{N-1} - \sigma'_N
\end{aligned}$$

If we combine all the above, we get a telescopic sum:

$$\sigma_1 - \sigma_N = \sigma'_1 - \sigma'_N$$

□

Proof. (Theorem 4.7) We will prove (4.16), and then (4.17) will follow due to symmetry. Let $(u, u'), (u, u'') \in \tilde{J}_k$ (i.e. connected nodes), such that $(u, u') \in \mathcal{U}_{\sigma_1,\sigma'_1}$ and $(u, u'') \in \mathcal{U}_{\sigma_2,\sigma'_2}$. We then have:

$$\begin{aligned}
u &= v_1 - \sigma_1 = v_2 - \sigma_2 \\
u' &= v_1 - \sigma'_1 \\
u'' &= v_2 - \sigma'_2
\end{aligned}$$

Subtract $u' - u''$ and substitute $v_1 - v_2$:

$$u' - u'' = v_1 - v_2 - (\sigma'_1 - \sigma'_2) = \sigma_1 - \sigma_2 - (\sigma'_1 - \sigma'_2) = 0$$

due to Lemma A.1, and $u' = u''$. □

A.4 Proofs for Chapter 5

Lemma 5.1

Proof. Since $\eta(\mathcal{E})$ is finite, we can re-write $R_{\mathcal{E}+\gamma}$ as a Haar integral:

$$\begin{aligned}\langle s|R_{\mathcal{E}}|t\rangle &= \frac{1}{\eta(\mathcal{E})}\int_{\mathcal{E}} x[s+\tau]x^*[t+\tau]d\eta(\tau) \\ &= \frac{1}{\eta(\mathcal{E})}\int_{\mathcal{E}+t} x[s-t+\tau]x^*[\tau]d\eta(\tau-t) \qquad \text{(Haar invariance)} \\ &= \frac{1}{\eta(\mathcal{E})}\int_{\mathcal{E}+t} x[s-t+\tau]x^*[\tau]d\eta(\tau) = (*)\end{aligned}$$

The integrand $f_{s-t}(\tau) = x[s-t+\tau]x^*[\tau]$ depends on $s-t$, alas, the integration domain depends on t solely. Recall however, that in general, for $A, B \subset \mathbb{T}$ we have

$$\int_B f d\eta = \int_A f d\eta + \left[\int_{B\setminus A} f d\eta - \int_{A\setminus B} f d\eta\right]$$

so that

$$(*) = \frac{1}{\eta(\mathcal{E})}\int_{\mathcal{E}} f_{s-t}(\tau)d\eta(\tau) + \epsilon(t) = R(s-t) + \epsilon(t)$$

with the residual

$$\epsilon(t) = \frac{1}{\eta(\mathcal{E})}\left[\int_{(\mathcal{E}+t)\setminus\mathcal{E}} f_{s-t}(\tau)d\eta - \int_{\mathcal{E}\setminus(\mathcal{E}+t)} f_{s-t}(\tau)d\eta\right]$$

If f_{s-t} is bounded by M, then we can bound the boundary residue by

$$|\epsilon(t)| \leq M\frac{\mu((\mathcal{E}+t)\Delta\mathcal{E})}{\mu(\mathcal{E})}$$

□

5.3

Proof. Let $K \subset \mathbb{T}$ be a finite collection of indices. The joint distribution of any finite sampling $\mathtt{D}_K$ satisfies

$$\begin{aligned}
\mathbf{p}(\mathtt{D}_{t+K} x_\tau \in A) &= \int_{\mathbb{T}} \mathbf{p}(\mathtt{D}_{t+K} \varsigma^\tau x \in A | \tau) d\mathbf{p}(\tau) \\
&= \int_{\mathbb{T}} \mathbf{p}(\mathtt{D}_K \varsigma^\tau x \in A | \tau) d\mathbf{p}(\tau - t) \qquad \text{(shift invariance)} \\
&= \int_{\mathbb{T}} \mathbf{p}(\mathtt{D}_K \varsigma^\tau x \in A | \tau) d\mathbf{p}(\tau) \\
&= \mathbf{p}(\mathtt{D}_K x_\tau \in A)
\end{aligned}$$

The deterministic case is treated by setting $\mathbf{p}(\mathtt{D}_K x \in A) = \mathbb{1}_A(\mathtt{D}_K x)$. □

Lemma 5.5

Proof. First note that $\psi(t)$ is strictly stationary due to Lemma 5.3, and so are $m(\psi(t)), \bar{m}(\psi(t))$, and $f(\psi(t))$. The two added components in $\mathcal{C}(b, f)$ are WSS, since

$$b(t) \cdot \bar{m}(\psi(t))$$

is a product of two *independent* WSS processes, and

$$f(\psi(t)) \cdot m(\psi(t))$$

is stationary, again by Lemma 5.3.

In order for their sum to be WSS, they need to be jointly WSS. Their cross-correlation can be written (due to the independence of b) as

$$\mathbb{E}\left\{b(t) \cdot \bar{m}(\psi(t)) f(\psi(s)) \cdot m(\psi(s))\right\} = \mathbb{E}(b(t))\mathbb{E}(f(\tau)),$$

where

$$f(\tau) = \bar{m}(\psi(t)) \cdot f(\psi(s)) \cdot m(\psi(s)) = \bar{m}(\phi(t+\tau)) \cdot f(\psi(s+\tau)) \cdot m(\psi(s+\tau))$$

Since $f(\tau)$ is stationary (which is again, are corollary of Lemma 5.3), its expectation is constant, and so overall $\mathcal{C}(f, b)$ is a sum of two jointly WSS processes and thus WSS on its own. □

Lemma 5.7

Proof. Let $Q \in \mathbb{R}^{D\times D}$ be a rotation matrix. Note that the probability measure of R is invariant under further rotations, $d\mathbf{p}(R) = d\mathbf{p}(RQ)$ for all rotations Q. Furthermore, under $\mathcal{B}(\mathbb{R}^D)$, we have associativity $(RQ)K = R(QK)$ (which is not true in general as rotation outside $\mathcal{B}(\mathbb{R}^D)$ might warp on the torus $[-1,1]^D$).

$$\begin{aligned}
\mathbf{p}(\mathtt{D}_{QK}x \in A) &= \mathbf{p}(\mathtt{D}_{R(QK)}y \in A) && (RQ)K = R(QK), \quad K \subset \mathcal{B}(\mathbb{R}^D) \\
&= \int \mathbf{p}(\mathtt{D}_{(RQ)K}y \in A|R)d\mathbf{p}(R) && \tilde{R} = RQ \\
&= \int \mathbf{p}(\mathtt{D}_{\tilde{R}K}y \in A|\tilde{R})d\mathbf{p}(\tilde{R}Q^*) && \text{(Rotation invariance)} \\
&= \int \mathbf{p}(\mathtt{D}_{\tilde{R}K}y \in A|\tilde{R})d\mathbf{p}(\tilde{R}) \\
&= \mathbf{p}(\mathtt{D}_{RK}y \in A) \\
&= \mathbf{p}(\mathtt{D}_{K}x \in A)
\end{aligned}$$

which completes the proof. □

A.4.1 Lemma 5.9

Proof. Note that $|y(t)| \leq M\mathbb{1}_B(t)$ where $\mathbb{1}_B(t)$ indicates $\mathcal{B}_{\mathcal{R}}([-1,1]^D)$, and $\mathbb{1}_B(Rt) = \mathbb{1}_B(t)$ for all R (since B is within the unit ball). The process $x(t) = y(R(t+\tau))$ is WSS with radial symmetry (due to Lemma 5.7) and

$$r(t) = \mathbb{E}(x(t)x(0)) = \int\int y(R(t+\tau))y(R\tau)d\mathbf{p}(\tau)d\mathbf{p}(R)$$

Note that

$$\begin{aligned}
|r(t)| &\leq \int\int |y(R(t+\tau))||y(R\tau)|d\mathbf{p}(\tau)d\mathbf{p}(R) \\
&\leq M^2\int\int \mathbb{1}_B(R(t+\tau))\mathbb{1}_B(R\tau)d\mathbf{p}(\tau)d\mathbf{p}(R) \qquad \mathbb{1}_B(Rt) = \mathbb{1}_B(t) \\
&= M^2\int\int \mathbb{1}_B(t+\tau)\mathbb{1}(\tau)d\mathbf{p}(\tau)d\mathbf{p}(R) \\
&= M^2\mathbf{p}(B\cap(B+t)) \\
&= 2M^2\beta_{\mathcal{R}}(\|t\|)
\end{aligned}$$

□

APPENDIX B

MATHEMATICAL PRIMER

B.1 Function Modules and Spaces

An important construction in signal processing is the module of functions from some set A to a ring R (commutative, with a unit, usually a field $\mathbf{k}$):

Definition B.1 (module of R-valued functions)**.** *The* function module *between A to R is the set comprised of all functions from A to R:*

$$R^A := \{f : A \to R\}$$

along with a natural R-linear structure of entry-wise linear combinations:

$$(\alpha f + \beta g)[u] := \alpha f[u] + \beta g[u]$$

where $f[u]$ denotes the value of f at index $u \in A$.

The dual module (see [26]) is $M^\vee$ the module of all R-linear functionals from M to R. Of course, whenever $R = \mathbf{k}$ is a field, then $\mathbf{k}^A$ is a vector space, and its dual will be denoted with an asterisk. Many results from linear algebra of vector spaces hold for modules over commutative rings.

We define the *impulse function* $\delta_{u_0} \in R^A$ as

$$\delta_{u'}[u] := \begin{cases} 1 & u = u' \\ 0 & \text{else} \end{cases} \tag{B.1}$$

The *valuation functional* $\delta_u^* x \in M^\vee$ is respective defined as

$$\delta_u^* x := x[u] \tag{B.2}$$

for every $x \in R^A$.

The *support* of a function $f \in R^A$ will be the space of all finitely supported

functions as:

$$\operatorname{supp}(f) = \{u \in A \ : \ f[u] \neq 0\} \tag{B.3}$$

$$\ell_0(R^A) = \{f \in R^A \ : \ \operatorname{supp}(f) \text{ is a finite set } \} \tag{B.4}$$

Clearly $\ell_0(R^A)$ is a submodule of R^A (every finitely supported function is a function). However, if $B \subset A$, the R^B is technically not a submodule of R^A, but naturally embedded by the so-called zero-padding $T : \mathbb{R}^B \hookrightarrow R^A$:

$$(Tv)[u] := \begin{cases} v[u] & u \in B \\ 0 & u \in A \setminus B \end{cases}.$$

The module R^A is an excellent tool to theoretically analyze R-values signals that are indexed by A, but rather impractical when A is infinite. This is where we result to

Definition B.2 (Generated Function Space)**.** *Fix some index set A. For every $B \subseteq A$ define*

$$\langle B\rangle := \left\{ \sum_{k=1}^{N} \alpha_k \delta_{u_k} \ \middle| \ \begin{array}{c} \alpha_1, \dots \alpha_N \in R \\ u_1, \dots u_N \in B \end{array}, \ \ N \geq 1 \right\} \subset R^A \tag{B.5}$$

Now we have $\langle B\rangle \subset \langle A\rangle$ whenever $B \subset A$, and also when B is a finite subset of A, every vector in $\langle B\rangle$ is a sum of finite basis elements. The drawback is loss of topological completeness (signal sequences may converge to limits outside $\langle B\rangle$). Note that for every $B \subseteq A$ we have

$$\ell_0(R^B) = \langle B\rangle \tag{B.6}$$

Also, whenever A is finite, then

$$\ell_0(R^A) = R^A \cong R^{|A|} \tag{B.7}$$

(here R^n is just the direct sum of R with itself n times). For example, if $A = \{a_1, \dots, a_n\}$ then $\langle A\rangle \cong R^n$, where the elements of A are usually indices for elements in $\langle A\rangle$.

An alternative notation (and mostly cosmetic) for δ_u and δ_u^* are the *Dirac notations*: for $u \in A$ define a corresponding *ket* vector and *bra* functional:

$$|u\rangle := \delta_u \in \langle A\rangle \qquad\qquad \langle u| := \delta_u^* \in \langle A\rangle^\vee \tag{B.8}$$

as defined above.

B.1.1 The Standard Inner and Scalar Products

The module $\langle A\rangle$ has a natural *standard* scalar product:

$$\langle f, g\rangle = \sum_{u\in A} f[u]g[u] \tag{B.9}$$

which is a bi-linear symmetric form over $\langle A\rangle \times \langle A\rangle$ (for $R = \mathbb{C}$ we modify the definition to an inner product, using $\overline{g[u]}$). Of course, the sum (B.9) is finite and well defined.

The Dirac notation system has been originally used in Hilbert spaces, where every vector $|u\rangle$ has a unique dual (functional), denoted $\langle u|$. Inner product is the concatenation:

$$\langle u|u'\rangle := |u'\rangle^* (|u\rangle) \tag{B.10}$$

Much of the elegance of the Dirac notation system carries on to general fields with scalar product, leaving out properties unique to inner products (i.e. positive definiteness). Nevertheless, the notion of *orthogonality* persists: two vectors $|u\rangle, |u'\rangle \in U$ are said to be orthogonal if $\langle u|u'\rangle = \delta_{u,u'}$, with the caveat that nonzero vectors can be orthogonal to themselves if $\mathrm{char}(\mathbf{k}) > 0$.

The standard basis of $\langle A\rangle$, which is $\{|u\rangle \mid u \in A\}$, is orthogonal by construction with respect to the (standard) scalar product:

$$\langle u|u'\rangle = \delta_{u,u'} := \begin{cases} 1 & u = u' \\ 0 & \text{else} \end{cases} \tag{B.11}$$

The *resolution of identity* is the decomposition of the identity map on $\langle A\rangle$:

$$I = \sum_{u\in A} |u\rangle \langle u| \tag{B.12}$$

B.1.2 Spaces of Linear Operators

Consider a pair (U, V) of finite dimensional linear spaces over $\mathbf{k}$. We denote

$$\mathtt{Lin}(U, V) := \mathrm{Hom}_{\mathbf{k}}(U, V) = V \otimes U^*$$

the (linear) space of $\mathbf{k}$-linear maps from U to V. For convenience, we denote the endomorphism space (self-mappings) by $\mathtt{Lin}(U) := \mathtt{Lin}(U, U)$.

Operators in $\mathtt{Lin}(U, V)$ are uniquely characterized by matrices, given fixed bases for U and V. Let $\mathcal{U}$ and $\mathcal{V}$ be two finite index sets for the bases $\{|u\rangle\}_{u\in\mathcal{U}}$ and $\{|v\rangle\}_{\in\mathcal{V}}$ of $U = \mathbf{k}^{\mathcal{U}}$, $V = \mathbf{k}^{\mathcal{V}}$.

Using the resolution of identity (B.12), we infer that every linear operator $S \in \mathtt{Lin}(U, V)$ can be represented in various ways by decomposing the identities on U an V:

$$S = \sum_{v,u} S_{v,u} \, |v\rangle \, \langle u| \qquad \text{bow-tie sum} \tag{B.13}$$

$$= \sum_{v\in\mathcal{V}} |v\rangle \cdot \langle v| \, S \qquad \text{covector sum} \tag{B.14}$$

$$= \sum_{u\in\mathcal{U}} S \, |u\rangle \cdot \langle u| \qquad \text{vector sum} \tag{B.15}$$

The matrix (rank 2 tensor) $S_{v,u} : \mathcal{V} \times \mathcal{U} \to \mathbf{k}$ given by

$$S_{v,u} = \langle v|S|u\rangle \tag{B.16}$$

is the so-called *matrix representation of* S with respect to the bases $|u\rangle$ and $|v\rangle$. The space of complex matrices (functions) indexed by $\mathcal{U} \times \mathcal{V}$ is isomorphic to $\mathtt{Lin}(U, V)$. The outer products $|v\rangle \, \langle u|$ define a standard basis of $\mathtt{Lin}(U, V)$ with respect to the bases $\{|u\rangle\}, \{|v\rangle\}$. From now on, we will use the terms *matrix* and *operator* interchangeably as the bases $|v\rangle$ and $|u\rangle$ are fixed.

The first thing we note about $\mathtt{Lin}(U, V)$ is it linear structure over $\mathbf{k}$, with the linear combinations defined by

$$(\alpha S_1 + \beta S_2) \, |u\rangle := \alpha S_1 \, |u\rangle + \beta S_2 \, |u\rangle$$

If $U \cong \mathbf{k}^s, V \cong \mathbf{k}^t$, there is a natural isomorphism

$$\mathtt{Lin}(U, V) \cong \langle \mathcal{V} \times \mathcal{U} \rangle \cong \mathbf{k}^{\mathcal{V}\times\mathcal{U}} \cong \mathbf{k}^{t\times s}$$

and of course, affine varieties are naturally defined on $\mathtt{Lin}(U, V)$ through polynomials in $\mathbf{k}[\mathcal{V} \times \mathcal{U}]$.

An operator $A \in \mathtt{Lin}(U, V)$ is said to be of *finite rank* if its range is a finite dimensional subspace of V (which is granted whenever V is finite

dimensional itself). In a Hilbert space and finite dimensional spaces, a finite rank operator of rank n has a normal form given by the finite sum

$$A = \sum_{i=1}^{n} |v_i\rangle \langle u_i| \tag{B.17}$$

B.2 Random Extensions of Deterministic Hilbert Spaces

Data samples live in a Hilbert space V, measurable with Borel sets naturally defined by open sets in V. Making samples in V random vectors is as simple as attaching a probability space $(\Omega, \mathcal{F}, \mathbf{p})$ to them: a *random data vector/process* in V is merely a measurable mapping $x : \Omega \to V$. Therefore $x(\omega) \in V$ is a *realization* vector. For the sake of consistency with previous notations, $x[t]$ will denote the random variable at a fixed index $t \in \mathbb{T}$ when $V = \mathbb{C}^{\mathbb{T}} \otimes \mathbb{C}^{L}$ leaving the random coordinate ω undetermined.

B.2.1 Moments and Geometry on Random Extensions

The notion of *expectation* (or sometimes called *mean*) is naturally defined on random vectors in V, as the mapping $\mathbb{E}$ of a process $x : \Omega \to V$ to a vector $\mathbb{E}(x) \in V$ defined by

$$\textbf{Expectation:} \qquad \mathbb{E}(x) := \int_{\Omega} x(\omega) d\mathbf{p}(\omega) \in V \tag{B.18}$$

The probabilistic extension holds for tensor builds over V. Every random vector x has an associate random co-vector x^* naturally defined on $u \in V$ by $(x^* u)(\omega) := x^*(\omega)u$. The space $V \otimes V^* \cong \mathtt{Lin}(V)$ extends by taking two processes $x, y : \Omega \to V$ into $(x \otimes y^*)(\omega) := x(\omega) \otimes y^*(\omega)$. This is merely an abstraction of the column-by-row matrix product, and is useful to define the *auto-correlation* operator as below:

$$\begin{aligned} \textbf{Auto-Correlation:} \qquad & R_x := \mathbb{E}(x \otimes x^*) \in \mathtt{Lin}(V) \\ \textbf{Cross-Correlation:} \qquad & R_{xy} := \mathbb{E}(x \otimes y^*) \in \mathtt{Lin}(V) \end{aligned}$$

The autocorrelation is self-adjoint and positive-semidefinite in the Hilbert context of V, so that all its eigenvalues are non-negative and its eigenvectors are orthogonal and span V.

A linear mapping $T : V \to U$, acting point-wise $(Tx)(\omega) := x(\omega)$ maps

the autocorrelation and mean as follows

$$\mu_{Tx} = T\mu_x \in U \qquad\qquad R_{Tx} = TR_xT^* \in \texttt{Lin}(U)$$

The probabilistic extension of V has similar geometry to V, provided by the inner product

$$\textbf{Inner-Product:} \qquad \langle x, y\rangle := \mathbb{E}\Big(\, \langle x(\omega), y(\omega)\rangle \,\Big) = \mathrm{tr}(R_{xy})$$

which in turn endows orthogonality, norm (hence distance), which is crucial for statistical analysis on V. We denote by $L_2(V)$ the collection of all random vectors of induced finite norm.

Laws of large numbers manifest the relation between statistical and probabilistic quantities. Given the samples $\{X_\tau\}_{\tau\in\mathcal{E}} \subset V$ drawn from the same distribution (independently, or with partial dependence), as $|\mathcal{E}| \to \infty$, under certain conditions we might have

$$\mu_{\mathcal{E}} := \frac{1}{|\mathcal{E}|}\sum_{\tau\in\mathcal{E}} X_\tau \to \mathbb{E}(X_\tau)$$
$$R_{\mathcal{E}} := \frac{1}{|\mathcal{E}|}\sum_{\tau\in\mathcal{E}} X_\tau \otimes X_\tau^* \to \mathbb{E}(X_\tau \otimes X_\tau^*)$$

with convergence depending on the context (mean squared convergence is perhaps the most convenient to work with in the Hilbert settings). The eigenvalue problem of $\mathbb{E}(X_\tau \otimes X_\tau^*)$ (known as the Karhunen-Loeve decomposition of random vectors) is the limit case of PCA.

B.2.2 Stationarity

For signals with spatial coordinate e.g. $V = \mathbb{C}^{\mathbb{T}} \otimes \mathbb{C}^L$, the expectation and auto-correlation tensors have explicit representations:

$$\mu_x[t] := \mathbb{E}(x[t]) \in \mathbb{C}^L$$
$$R_x[t,s] := \mathbb{E}(x[t] \otimes x[s]^*) \in \texttt{Lin}(\mathbb{C}^L)$$

Note: For a multichannel signal $R_x[t,s]$ is a matrix rather than a scalar.

The group structure of $\mathbb{T}$ gives meaning to the index difference $s-t$, and the notion of stationarity extends directly from the classical one-dimensional case:

Definition B.3 (Stationary)**.** *A process x on $\mathbb{T}$ is said to be*

1. Stationary *if the joint probability measure is shift invariant, i.e. for every finite Γ and every $\tau \in \mathbb{T}$ one has $\mathbf{p}(\mathtt{D}_\Gamma x) = \mathbf{p}(\mathtt{D}_\Gamma \varsigma^\tau x)$*

2. Wide sense stationary *(WSS) if its first and second moments are shift-invariant, that is, for all $\tau \in \mathbb{T}$ one has*

$$\mathbb{E}(\varsigma^\tau x) = \mathbb{E}(x) \qquad\qquad R_x \varsigma^\tau = \varsigma^\tau R_x$$

Stationary implies WSS, but not vice versa:

$$\mathbb{E}(\varsigma^\tau x) = \int_\Omega \varsigma^\tau x d\mathbf{p}(x) = \int_\Omega x d\mathbf{p}(\varsigma^{-\tau} x) = \int_\Omega x d\mathbf{p}(x) = \mathbb{E}(x)$$

The proof for the autocorrelation is similar.

B.3 Polynomial Algebra

In this section we formalize the concept of a polynomial mapping acting between free vector spaces, which generalized multivariate polynomials.

Let $\Sigma = \{\sigma_1, \ldots, \sigma_n\}$ be a finite set of indices, and let

$$X = \{X_\sigma\}_{\sigma \in \Sigma}$$

be a finite set of formal variables associated with Σ.

Definition B.4 (Polynomial Ring)**.** *The multivariate polynomial ring $\mathbf{k}[\Sigma]$ with coefficients in $\mathbf{k}$ is defined as the collection of finite formal sums:*

$$\mathbf{k}[X] := \left\{ \sum_\alpha c_\alpha \prod_{\sigma \in \Sigma} X_\sigma^{\alpha_\sigma} \quad , c_\alpha \in \mathbf{k} \right\} \tag{B.19}$$

where $\alpha : \Sigma \to \mathbb{N}$ are the corresponding powers (assuming commutativity between the elements of X), and by convention $X_\sigma^0 = 1$, the unit of $\mathbf{k}$.

Every polynomial $p \in \mathbf{k}[X]$ defines a polynomial mapping

$$p : \mathbf{k}^\Sigma \to \mathbf{k}$$

obtained by plugging the values of the input in $\mathbf{k}^\Sigma$ into their corresponding generators x_σ.

Traditionally, the set Σ are the ordinal integers (such that $\mathbf{k}^\Sigma$ is a Cartesian space), but this is by no means a requirement - polynomial maps can be natively defined on $\mathbf{k}^\Sigma$ to $\mathbf{k}$ without resorting to Cartesian space $\mathbf{k}^{|\Sigma|}$.

The definition in (B.19) seamlessly extends to vector-valued coefficients

Definition B.5 (Modules of Polynomial)**.** *Let M be a finite dimensional vector space over $\mathbf{k}$; we define the set*

$$M[X] := \left\{ \sum_\alpha c_\alpha \prod_{\sigma\in\Sigma} x_\sigma^{\alpha_\sigma} \quad , c_\alpha \in M \right\}$$

which is a linear space over $\mathbf{k}$ and a module of $\mathbf{k}[X]$.

Whenever M is an *algebra* (for example - endomorphisms $\mathtt{Lin}(U,U)$), then $M[X]$ is a module over M.

By abuse of notations we will write $\mathbf{k}[\Sigma]$ and $M[\Sigma]$ rather than $\mathbf{k}[X]$ and $M[X]$, carrying every index $\sigma \in \Sigma$ is carries an appropriate generator X_σ.

There is a canonical isomorphism

$$\mathcal{M}[X] \cong \mathbf{k}[X] \otimes M \tag{B.20}$$

(see [27]), which in simple words, means that we can think of elements of $M[X]$ both as "tensor of polynomials" as well as "polynomials with tensor coefficients".

Polynomial mappings are well defined between linear spaces, even without a specified basis.

Definition B.6 (Polynomial Mapping)**.** *Let $U = \mathbf{k}^\Sigma$ and $V = \mathbf{k}^\mathcal{V}$ be two linear finite dimensional spaces.*

A mapping $p : U \to V$ is said to be polynomial if the coefficient $\langle v|p(\theta)|$ is polynomials in the coefficients $\{\langle\sigma|\theta|\}$. This characterization does not alter under change of basis.

B.3.1 Affine and Homogeneous Varieties

Affine varieties in $\mathbf{k}^\Sigma$ correspond to the zero sets of collections of polynomials in $\mathbf{k}[X]$. We can *"pack"* a collection of K polynomials to a mapping in the module $\mathbf{k}^K[X]$ (or any K-dimensional vector space).

Definition B.7 (Affine Variety)**.** *For a polynomial mapping $p \in M[X]$ then*

$$\mathbb{V}(p) := \{\theta \in \mathbf{k}^{\Sigma} \ : \ p(\theta) = 0_{\mathcal{M}}\} \tag{B.21}$$

Another important concept is homogeneity:

Definition B.8 (Homogeneous Polynomial and Variety)**.** *We say that $p \in \mathcal{M}[X]$ is* homogeneous *if there exists some integer $d > 0$ such that*

$$p(\lambda\theta) = \lambda^d p(\theta) \quad \textit{for all} \quad \theta \in \mathbf{k}^{\Sigma}, \quad \lambda \in \mathbf{k}$$

An affine variety $\mathcal{M} = \mathbb{V}(p)$ is said to be homogeneous *if p is a homogeneous polynomial, or equivalently*

$$\lambda\mathcal{M} \subseteq \mathcal{M}$$

for all $\lambda \in \mathbf{k}$.

Definition B.9 (Coordinate Ring)**.** *For a given affine variety $V \in \mathbf{k}^D$, the coordinate ring is the quotient ring*

$$\mathbf{k}[V] := \mathbf{k}[x_1, \ldots, x_d]/\mathcal{I}(V)$$

Alternatively, it can be thought of as the set of all $\mathbf{k}$-valued polynomial functions defined over V.

Definition B.10. *An algebraic variety V is said to be* reducible *if it can be represented as a union of two other varieties.*

Algebraic varieties define a topology in $\mathbf{k}^D$ by closed sets, known as Zariski topology: a set $X \subset \mathbf{k}^D$ is closed if X is an affine variety (hence X is open if its complement is a variety).

B.4 Grassmanians

For the analysis of range-space based identification, we will need to appeal to a structure containing all linear spaces of a given dimension, known as a *Grassman set* (or a *Grassmanian*).

Let U be a n-dimensional vector space over $\mathbf{k}$. The set of all r-dimensional linear subspaces U is known as the *Grassmanian* (see [28] for more details):

$$\mathtt{Gr}(U, r) := \{V \subset U \ : \ \dim(V) = r\}$$

The set $\mathtt{Gr}(U, r)$ can be given several different structures: a topological space (metric space in fact, and even a smooth manifold), and affine/projective algebraic variety.

B.4.1 Grassmanian as an Algebraic Variety

The Grassmanian possesses a structure of an algebraic variety - a nullset of polynomials, given by the renowned Plücker embedding. The idea is to map any arbitrary basis to its the exterior product:

$$V \mapsto \{v_1, \ldots, v_r\} \mapsto v_1 \wedge v_2 \wedge \cdots \wedge v_r$$

For any other basis of V, the wedge product would alter only by a scalar, and in particular

$$\tilde{v}_k = \sum_{m=1}^{n} a_{mk} v_m$$

then

$$\tilde{v}_1 \wedge \cdots \wedge \tilde{v}_s = \det(A) v_1 \wedge \cdots \wedge v_r$$

Here $A = (a_{mk})$ is the non-signular square basis-change matrix, so that $\mathrm{span}_k\{v_k\}$ is identified with the wedge product of its various bases (up to a scalar).

The embedding $\psi : \mathtt{Gr}(U, r) \hookrightarrow \mathbb{P}(\bigwedge^r U)$ is well defined by

$$\psi(V) := [v_1 \wedge v_2 \wedge \cdots \wedge v_r] \tag{B.22}$$

where $\{v_1, \ldots, v_s\}$ is an arbitrary basis of V, and the square bracket denotes to the projective class in the projectivized wedge product.

B.4.2 Grassmanian as a Smooth Topological Manifold

Topology on the Grassmanian enables concepts such as continuity and separability. Whenever $\mathbf{k} = \mathbb{R}$ or $\mathbf{k} = \mathbb{C}$, we can equip $\mathtt{Gr}(U, r)$ with a metric (making it a topological space thereof), for example, for $V_1, V_2 \in \mathtt{Gr}(U, r)$ we define their distance as:

$$d(V_1, V_2) := \sup_{v_1 \in B_1} \inf_{v_2 \in V_2} \|v_2 - v_1\| \tag{B.23}$$

where $\|\cdot\|$ is the standard Cartesian norm on U, and B_1 is the unit sphere of V_1 with regard to this norm.

Furthermore, under this assumption, the Grassmanian is also a smooth manifold (as any projective variety is) of dimension $r(t-r)$

REFERENCES

[1] C. R. Johnson, H. Šmigoc, and D. Yang, "Solution theory for systems of bilinear equations," *Linear and Multilinear Algebra*, vol. 62, no. 12, pp. 1553–1566, 2014.

[2] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *Signal Processing, IEEE Transactions on*, vol. 43, no. 2, pp. 516–525, 1995.

[3] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *Information Theory, IEEE Transactions on*, 1994.

[4] G. Harikumar and Y. Bresler, "Channel order estimation for multichannel blind image recovery," in *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No.98CH36284)*, vol. 2, Nov 1998, pp. 984–988 vol.2.

[5] G. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms," *Image Processing, IEEE Transactions on*, vol. 8, no. 2, pp. 202–219, 1999.

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, July 2006. [Online]. Available: http://doi.acm.org/10.1145/1141911.1141956

[7] F. Šroubek, G. Cristóbal, and J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution," in *IEEE Tran. Imag. Proc*, vol. 2, 2007.

[8] F. Király and R. Tomioka, "A combinatorial algebraic approach for the identifiability of low-rank matrix completion," *arXiv preprint arXiv:1206.6470*, 2012.

[9] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *arXiv preprint arXiv:1211.5608*, 2012.

[10] S. Choudhary and U. Mitra, "On identifiability in bilinear inverse problems," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4325–4329.

[11] D. L. Pimentel-Alarcón, N. Boston, and R. D. Nowak, "Deterministic conditions for subspace identifiability from incomplete sampling," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 2191–2195.

[12] S. Choudhary and U. Mitra, "Sparse blind deconvolution: What cannot be done," in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 3002–3006.

[13] J. H. Manton and W. D. Neumann, "Totally blind channel identification by exploiting guard intervals," *Systems & Control Letters*, vol. 48, no. 2, pp. 113–119, 2003.

[14] J. H. Manton, W. D. Neumann, and P. T. Norbury, "On the algebraic identifiability of finite impulse response channels driven by linearly precoded signals," *Systems & Control Letters*, vol. 54, no. 2, pp. 125–134, 2005.

[15] L. Ljung, *System Identification*. Springer, 1998.

[16] M. Gevers et al., "System identification without Lennart Ljung: What would have been different?" *Forever Ljung in System Identification, Studentlitteratur AB, Norrtalje*, 2006.

[17] E. Yarkony, Y. Baryshnikov, and Y. Bresler, "Range-space based identification of parametric linear systems," in *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, July 2016, pp. 1–5.

[18] D. Mumford, *Algebraic Geometry: Complex Projective Varieties*. Springer, 1995.

[19] D. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*. Springer, 2006.

[20] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[21] G. Templeton, "World's most powerful MRI can lift a tank like magneto, or see deep into your brain," 2013. [Online]. Available: https://www.extremetech.com/extreme/169526-worlds-most-powerful-mri-can-lift-a-tank-like-magneto-or-see-deep-into-your-brain

[22] L. Beilina, E. Karchevskii, and M. Karchevskii, *Numerical Linear Algebra: Theory and Applications*. Springer International Publishing, 2017. [Online]. Available: https://books.google.com/books?id=RijZAQAACAAJ

[23] A. Deitmar and S. Echterhoff, *Principles of Harmonic Analysis*, ser. Universitext. Springer New York, 2008. [Online]. Available: https://books.google.com/books?id=w40NcyUM8iwC

[24] D. Zoran and Y. Weiss, "Natural images, Gaussian mixtures and dead leaves," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1736–1744. [Online]. Available: http://papers.nips.cc/paper/4758-natural-images-gaussian-mixtures-and-dead-leaves.pdf

[25] M. Unser, "On the approximation of the discrete Karhunen-Loeve transform for stationary processes," *Signal Processing*, vol. 7, no. 3, pp. 231 – 249, 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0165168484900021

[26] S. Lang, *Algebra*, ser. Graduate Texts in Mathematics. Springer New York, 2005. [Online]. Available: https://books.google.com/books?id=Fge-BwqhqIYC

[27] M. Atiyah, *Introduction To Commutative Algebra*, ser. Addison-Wesley Series in Mathematics. Westview Press, 1994. [Online]. Available: https://books.google.com/books?id=HOASFid4x18C

[28] J. W. Milnor and J. D. Stasheff, *Characteristic Classes*. Princeton University Press, 1974, no. 76.